THE UNIVERSITY
of EDINBURGH

Mapping run-of-river hydropower resource of large catchments

Antony David Walker

A thesis submitted for the degree of Doctor of Philosophy.

The University of Edinburgh.

September 2018

Abstract

There is overwhelming scientific evidence that shows the temperature of the Earth's atmosphere is rising at an unprecedented rate. This is attributed to increased levels of greenhouse gas emissions, a large proportion of which originates from anthropogenic combustion of carbon-based fossil fuels for energy. There is therefore a strong argument for the increased role of less environmentally damaging, low carbon energy sources including renewable energy technologies. Run-of-river hydropower is one such renewable energy option, considered more environmentally benign than traditional hydropower which requires the construction of large dams to create a reservoir.

The aim of this study was to develop a model to search for, and map, economically viable run-of-river hydropower resource that can function on any global catchment of any size. Development and testing of the model was conducted on China's 2 million km^2 Yangtze River drainage basin, the third longest river in the world and a rich landscape for hydropower.

A gridded, distributed hydrological model was developed integrating high-resolution meteorological datasets and a digital elevation model (DEM). Using the model, the surface hydrology of the Yangtze catchment was simulated at a timestep of 6 minutes to obtain the mean daily surface runoff for every day from the beginning of 1979 to the end of 2007. Observed river flow data from sub-catchments of the Yangtze were used to calibrate the model by differential optimisation, an evolutionary computation technique. Validation was carried out on a 1.6 million km^2 sub-catchment resulting in a mean objective function of 0.95 (where a perfect fit would be 1.0) across 8 objective functions commonly used in hydrology.

Catchment wide mean daily runoff data was used to develop flow duration curves across the catchment river network. Virtual power stations were constructed at each river cell, iteratively testing differing scheme configurations, and costed using the RETScreen methodology. A best performing hydropower network was determined by a conflict algorithm, designed to prioritise high profit schemes and to remove lower performing and conflicting schemes. This resulted in a potential run-of-river installed capacity across the Yangtze catchment of 103GW (at 10% discount rate), generating 394TWh per annum. This model would be a valuable tool in finding optimal locations for future hydropower resource.

Declaration of originality

I hereby declare that the research recorded in this thesis and the thesis itself was composed and originated entirely by myself in the School of Engineering at The University of Edinburgh.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Partnership.

I would like to thank professors Tom Bruce and Gareth Harrison for their supervision, expertise and positivity, and Professor Robin Wallace and Dr Gerald Muller for their very thorough examination. Particular thanks go to Professor Clive Greated for encouraging me to take on a PhD and for his general support, who sadly died shortly after the viva of this PhD .

Many thanks to Neil Gourlay, Managing Director of the Auchenchyne farm, for his warm welcome and enthusiasm during the visit to his micro-hydropower installations. Similarly, Gordon Black and Nick Forrest from Babyhydro during their installation of a micro-hydro system at Dawyck Botanic Gardens, and Richard Thorne for showing us around the community hydro scheme in Mull. Massive thanks for the time and effort of Mark and Angela from EDDIE support, who helped me navigate the compute cluster. Thanks to David Williams and others who provided advice during my trip to the Hydropower and Dams conference in Laos.

I will be ever grateful to Clive, Nicola Gibbs, Richard Adlington and Reverend Dr. Andrew Knight for helping proof read the thesis.

Thanks to my wife Alicia for further proof reading and for reminding me it will all be worthwhile. To my wonderful children, Yannik and Orlana (Orlana was born one month after starting this Ph.D) and our dog Momo, for constantly reminding me not to work too much. Finally, to Mum and Dad for teaching me to have an open mind and not to be frightened of trying new things.

In memory of Dad, Piers Walker, 1942 – 2014 and Professor Clive Greated, 1940 – 2018.

Contents

List of figures

List of tables

Acronyms and abbreviations

2D	Two-dimensional
AGCM	Atmosphere General Circulation Model
APHRODITE	Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
ATI	Antecedent temperature index
BP	British Petroleum
CGIAR-CSI	Consultative Group for International Agricultural Research Consortium for Spatial Information
CCS	Carbon capture and storage
CEH	Centre of Ecology and Hydrology
CIAT	International Center for Tropical Agriculture
CMC	Canadian Meteorological Centre
CNY	Chinese Yuan
CRAN	Comprehensive R Archive Network
CRU	Climatic Research Unit
CSFR	Climate Forecast System Reanalysis
D∞	D infinity
D8	Deterministic 8 directions
DCW	Digital Chart of the World
DEM	Digital elevation model
DEMON	Digital elevation model networks
DEoptim	Differential evolution optimisation, R package
DIAS	Data Integration and Analysis System
DSM	Digital surface model
DTED	Digital Terrain Elevation Data
DTM	Digital terrain model

EC	Efficiency criteria
EDCF	Edinburgh (University) Compute and Data Facility
EDDIE	Edinburgh University Compute and Data Facility - nickname
EFAS	European Flood Awareness System
ESRI	Environmental Systems Research Institute
ECMWF	European Centre for Medium-Range Weather Forecasts
FAR15s	Flow accumulation raster at 15-arcseconds raster
FAO	Food and Agriculture Organization of the United Nations
FDC	Flow duration curves
FDR	Flow direction raster
FNN	Fast Nearest Neighbour Search Algorithms
G2G	Grid-to-grid hydrological model
GDP	Gross Domestic Product
GEOS DAS	Goddard Earth Observing System Data Assimilation System
GEOTIFF	Georeferenced Tagged Image File Format
GLUE	General Likelihood Uncertainty Estimation
GLWD	Global Lakes and Wetlands Database
GIS	Geographic information system
GMTED2010	Global Multi-resolution Terrain Elevation Data 2010
GRDC	Global Runoff Data Centre
GRP	Glass fibre reinforced plastics
GSI	Grid-point Statistical Interpolation
GTOPO30	Global 30-Arcsecond Elevation DEM
GTS	Global Telecommunications System
HDEM15s	HydroSHEDS digital elevation model at 15-arcseconds resolution
HFDR15s	HydroSHEDS flow direction raster at 15-arcseconds resolution
HydroSHEDS	Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales
HYDRO1k	A geographic database developed to provide comprehensive and consistent global coverage of topographically derived datasets

IAU	Incremental Analysis Update
IDW	Inverse distance weighting
IPCC	Intergovernmental Panel on Climate Change
IRENA	International Renewable Energy Agency
JPL	Jet Propulsion Laboratory
Landsat 5	Low Earth orbit satellite launched on March 1, 1984 to collect imagery of the surface of Earth
LCOE	Levelised cost of electricity
MCMC	Markov Chain Monte Carlo
MEaSUREs	Making Earth System Data Records for Use in Research Environments
MERRA	Modern Era Retrospective-Analysis for Research Applications
MFD	Multiple Flow Direction algorithm
MODIS	Moderate Resolution Imaging Spectroradiometer
mHM	mesoscale Hydrological Model
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NetCDF	Network Common Data Form
NGA	National Geospatial-Intelligence Agency
NOAA	National Oceanic and Atmospheric Administration
NPV	Net present value
NTSG	Numerical Terradynamics Simulation Group
OECD	Organisation for Economic Co-operation and Development
PERSIANN-CDR	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Climate Data Record
PERSIANN-CCS	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Cloud Classification System
R	R programming language for statistical computing and graphics
Rcpp	R to C++, R package
RGDAL	Bindings for the Geospatial Data Abstraction Library

RGEOS	Interface to Geometry Engine – Open Source
RMB	Chinese Yuan Renmibi
RMSE	Root mean square error
ROR	Run-of-river
rstn	Ratio of 0.05° grids with station [%]
SAR	Synthetic aperture radar
SRTM	Shuttle Radar Topography Mission
SWAT	Soil and Water Assessment Tool
SWDB	SRTM water body dataset
SWE	Snow water equivalent
TERRA	Meaning Earth - multi-national NASA scientific research satellite in a Sun-synchronous orbit around the Earth
TIFF	Tagged Image File Format
TS	Surface temperature
TOA	Top of the atmosphere
UCI	University of California
USGS	United States Geological Survey
WGS84	World Geodetic System 1984
WORLDCLIM	World Climatology Data
WWF	World Wide Fund for Nature
YDB	Yangtze drainage basin
YDB15s	Yangtze drainage basin at 15-arcseconds

Nomenclature

$\bar{O}$	mean observed runoff [m^3s^{-1}]
O_i	observed runoff values at each FDC percentile [m^3s^{-1}]
$\bar{P}$	mean predicted runoff [m^3s^{-1}]
$\bar{T}$	mean daily temperature [°C]
$\bar{g}$	average topographic gradient
η_e	electrical efficiency %
η_t	efficiency turbine %
Δt_p	time interval of precipitation data [hours]
Δx	size of the grid square [m]
°C	degrees-Celsius
AC	alternating current
ATI_1	antecedent temperature index at previous timestep [°C]
ATI_2	antecedent temperature index at current timestep [°C]
A_v	seasonal variation adjustment (set to 1.0 for latitudes <54°N)
$_b$	denotes sub surface flows
b	parameter determined from grid square slope values
c	kinematic wave speed [ms^{-1}]
cal	calories
CH_4	methane
c_i	specific heat of ice [cal gm^{-1} $°C^{-1}$] = 0.5
c_{max}	regional upper limit of storage capacity
CO_2	carbon dioxide
c_p	specific heat of the air [$MJkg^{-1}°C^{-1}$]
CR	capillary rise [mm]
Ct	critical moisture value [mm]
D	degree day factor [$mm°C^{-1}day^{-1}$]
d	Drainage from the base of the grid within the timestep [mm]
d	index of agreement
d	zero plane displacement height [m]
D_2	heat deficit at current timestep [mm]
D_2	heat deficit at previous timestep [mm]
DAYGM	ground melt parameter (model parameter) [mm day^{-1}] = 0.3

DC	direct current
DDF_{ice}	degree day factor of ice, set to 7.1 [mm of water $°C^{-1}$ day^{-1}]
DP	deep percolation [mm]
E	Nash-Sutcliffe efficiency
E	excess liquid water [mm]
e(T)	vapour pressure at air temperature [Pa]
$e(T_D)$	vapour pressure at dew point [Pa]
e°(T)	saturation vapour pressure at the air temperature T [kPa]
E_1	average hourly lagged excess water available in time-period [mm]
E_{1s}	average hourly lagged excess liquid water available for Δ [inches]
E_{1s}	E_l / 25.4 (to convert mm to inches)
e_a	actual vapour pressure [kPa]
$e_s - e_a$	vapour pressure deficit [kPa]
e_s	mean saturation pressure [kPa]
ET_o	reference crop evapotranspiration [mm day^{-1}]
exp [..]	2.7183 (base of natural logarithm) raised to the power [..]
FRSAT	MERRA fractional saturated area (a fraction)
f_{snow}	fraction of snow in the mixed rain/snow
g	Acceleration due to gravity = 9.8067 [ms^{-2}]
G	soil heat flux [$MJm^{-2}day^{-1}$]
GB	giga-bytes
g_{max}	regional upper limit of topographic gradient
GW	gigawatt
h	DEM elevation [m]
h	crop height [m]
H	sensible heat [Wm^{-2}]
H_e	height elevation [m]
H_{net}	net head [m]
hr	hours
I	irrigation [mm]
$_k$	denotes positions in discrete space
k	von Karman's constant = 0.41
K	Kelvin
k_c	crop coefficient [dimensionless]
k_d	storage rate constant [mm^{-2}]

km	kilometre
kPa	kilo-Pascals
kV	kilo-volts
$_l$	denotes land
LAI	leaf area index [m^2m^{-2}]
L_f	latent heat of fusion [cal gm^{-1}] = 80
ls^{-1}	litres per second
LW	longwave radiation [MJm^{-2}]
M	molar mass of dry air = 0.0289644 [kg mol^{-1}]
m	metre [m]
MB	mega-byte
mb	millibar
MBASE	base temperature [°C] (model parameter) = 0°C
M_{dry}	molar mass of dry air [$kgmol^{-1}$] = 0.0289644
M_f	melt factor [$mm°C^{-1}\Delta t^{-1}$]
MFMAX	maximum melt factor – June 21st [$mm°C^{-1}6h^{-1}$] (model parameter)
MFMIN	minimum melt factor – December 21st [$mm°C^{-1}6h^{-1}$] (model parameter)
M_g	amount of ground melt during each time-period [mm]
M_{H2O}	molar mass of water [$kgmol^{-1}$] = 0.0181534
min	minute
MJ	mega-joules
mm	millimetres
M_{nr}	melt during non-rain periods [mm]
M_r	melt during rain-on-snow time intervals [mm]
MW	mega-watt
MWh	mega-watt hour
N	number of days since March 21st
N_2O	nitrous oxide
NMF	negative melt factor [$mm°C^{-1}6hr^{-1}$] (model parameter)
O_g	outflow due to ground melt for each time interval (mm)
O_{mr}	snow cover outflow from melt or rain-on-snow [mm]
O_{mr1}	hourly snow cover outflow from melt or rain-on-snow [mm]
O_s	total snow cover outflow for each precipitation data interval [mm]
P	atmospheric pressure [kPa]

P	electrical power [W]
P	rainfall / precipitation [mm]
P_0	atmospheric pressure at sea-level = 101325 [Pascals]
PET	potential evapotranspiration [mm day^{-1}]
P_{H20}	partial pressure of water [Pa]
P_i	predicted runoff values at each FDC percentile [m^3s^{-1}]
PLWHC	percent liquid water holding capacity (model parameter
P_r	precipitation in the form of rain [mm]
PWhr	petawatt hour ($1*10^{15}$ Whr)
q	direct runoff rate [mm]
q	channel flow
Q	flow [m^3s^{-1}] or [ls^{-1}]
Q_0	100^{th} percentile flow – flow exceeded 0% of the time
Q_{100}	0^{th} percentile flow – flow exceeded 100% of the time
Q_5	95^{th} percentile flow – flow exceeded 5% of the time
Q_w	liquid water available at the snow surface [mm]
r	denotes river
R	earth's radius (6371km)
R	return flow
R_0	universal gas constant = 8.31447 [J $mol^{-1}K^{-1}$]
R_1	one-hour withdrawal rate [mm]
r^2	coefficient of determination
r_a	bulk aerodynamic resistance [sm^{-1}]
$R_{absorbedl}$	absorbed longwave radiation at the surface [$MJm^{-2}day^{-1}$]
$R_{emittedl}$	emitted longwave radiation at the surface [$MJm^{-2}day^{-1}$]
RH_{max}	daily maximum relative humidity [%]
RH_{min}	daily minimum relative humidity [%]
R_{ice}	refreeze coefficient of ice, set to 0.05 [mm of water $°C^{-1}$ day^{-1}]
R_l	return flow land
R_r	return flow river
R_n	net radiation [$MJm^{-2}day^{-1}$]
R_{nl}	net outgoing LW radiation at top of the atmosphere [$MJm^{-2}day^{-1}$]
R_{ns}	net incoming SW radiation at the surface [$MJm^{-2}day^{-1}$]
RO	surface runoff [mm]
R_r	return flow river

r_s	bulk surface resistance [sm^{-1}]
s	seconds
S	water storage in grid cell [mm]
S_1	amount of lagged excess liquid water in storage at beginning of hour [mm]
S_2	storage at the end of the hour [mm]
SF	subsurface flow [mm]
SH	specific humidity [kg kg^{-1}]
S_{max}	maximum storage capacity [mm]
S_{new}	new snowfall amount [mm]
SW	shortwave radiation [MJm^{-2}]
SW	soil water content [mm]
T (or T_a)	air temperature [°C]
T_0	sea level standard temperature = 288.15 [K]
TB	terra-bytes
TIMP	model parameter = 0.1
T_{kv}	virtual temperature [˚C]
T_{new}	temperature of new snow [°C]
T_r	temperature of rain [˚C] = T_a or 0˚C (whichever greater)
T_{sur}	snow surface temperature [°C]
TWh	terawatt hours
u	lateral flow
UADJ	average wind function [mm mb^{-1} $6hr^{-1}$] (model parameter) = 0.04
u_z	wind speed at height z [ms-1]
W	watt
W_i	water equivalent of the ice portion of the snow cover [mm]
W_{is}	W_i / 25.4 (to convert mm to inches)
W_{melt}	snowmelt [mm]
W_q	ground melt parameter [mm day^{-1}] = 0.3
W_q	liquid water held by the snow [mm]
W_{qx}	liquid water capacity [mm]
x_{H2O}	mole fraction, volume mixing ratio of water
yr	year
z_h	height of humidity measurements [m]
z_m	height of wind resistance [m]
z_{oh}	roughness length governing transfer of heat and vapour [m]

z_{om}	roughness length governing momentum transfer [m]
α	surface albedo [fraction]
γ	psychometric constant [kPa^{o}C^{-1}]
Δ	slope of saturation vapour pressure temperature relationship [kPa^{o}C-1]
ΔD_p	change in the heat deficit due to snowfall [mm]
ΔD_t	change in heat deficit due to a temperature gradient [mm]
Δt	time-period [h]
ε	ratio of molecular weight of vapour/dry air = 0.622
λ	latent heat of vapourisation = 2.45 MJ kg^{-1} (at 20^{o}C)
λ ET	evapotranspiration fraction
ρ	density [kg m^{-3}]
ρ_a	mean air density [kgm^{-3}]
σ	Stefan-Boltzman constant – 6.12x10^{-10} [mm˚K^{-1}hr^{-1}]
φ	latitude
β	drainage parameter [dimensionless]
θ	the dimensionless wave speed $c\frac{\Delta t}{\Delta x}$ and $0 < \theta < 1$
κ	rate constant [s^{-1}]
λ	longitude

Hydropower costing nomenclature:

$\hat{}e_{nq}$	peak efficiency adjustment
e_p	peak efficiency
A	access road difficulty factor (set to 2 for all roads)
A	cross sectional area of penstock [m^2]
B	civil works 'B' factor
C	civil cost factor depending if an existing dam is in place
C_0	total investment costs [CNY]
C_{con}	combined includes exchange rate and the rate of inflation
C_{ex}	exchange rate
C_g	lower cost motor factor
C_{inf}	Chinese rate of inflation
C_t	net cash inflow during the period t
d	runner diameter [m]
D	transmission line installation difficulty factor (set to 1.5)
d_p	diameter of penstock [m]

d_t	turbine runner throat diameter [m]
E	annual energy produced [MWh]
E	engineering cost factor to account for existing dams in projects
E_c	local vs Canadian equipment cost ratio = 1.0
e_g	generator efficiency set to 0.98
e_q	efficiency at flows below peak
e_r	efficiency at full load
E_t	energy generation in year t (kWh)
e_t	turbine efficiency;
e_{tQd}	turbine efficiency at design flow
f	Darcy friction factor
F	frost factor
f	number of frost days
F_c	local vs Canadian fuel cost ratio = 0.94
F_t	fuel costs in the year t – set to zero for hydropower
G	grid connection factor
hf	head-loss due to friction [m]
hf_n	head-loss at a flow less than the design flow [m]
H_g	grosshead
h_{hydr}	hydraulic losses
h_{tail}	head-loss due to the tail water effect
I	income per MWh
i	interest rate (set to 6) [%]
i_{inf},	inflation rate
IQR	interquartile range
IRR	internal rate of return
I_t	capital costs in the year t
J_t	higher cost vertical axis turbine factor
K_t	lower cost small horizontal axis turbine factor
kW	kilo-watt
l_a	length of access road [km]
l_b	distance to borrow pits (set to 0.5km for all sites) [km]
L_c	local vs Canadian labour cost ratio = 0.90
l_d	dam crest length [m], where W = $7.2Q_5{}^{0.5}$
l_{dt}	percentage loss due to downtime (e.g. maintenance etc), set to 2%

$l_{hydr,max}$	maximum hydraulic head-loss percentage
l_p	length of penstock [m]
L_p	penstock length [m]
l_{para}	parasitic losses = 2%
l_T	length of transmission line [km]
l_{trans}	transformer losses = 2%
M_t	maintenance and operation costs in the year t
MWh	mega-watt hour
MW_u	capacity per turbine [MW]
n	Manning number for GRP pipe (0.009);
n	number of turbines
n	number of years = 25 [years];
n_p	number of penstocks
n_q	specific speed based on flow
O&M	operation and maintenance costs.
P	factor to reflect cost of wood pole vs steel tower construction
P	plant capacity [kW]
P_n	power at percentile flow Q_n
Q_d	design flow [m^3s^{-1}]
Q_{mean}	mean flow [m^3s^{-1}]
r	discount rate [%]
R	rock factor
R_e	Reynolds number
t	time-period [years]
T	tote road factor (set to 0.25 for all projects)
t_{av}	average thickness of penstock walls [mm]
t_b	thickness at base of penstock [mm]
T_{local}	a local installation factor
t_t	thickness at top of penstock [mm]
V	average water velocity [ms^{-1}]
v	kinematic viscosity of water [$1.31*10^{-6}m^2s^{-1}$]
V	transmission voltage [kV]
V_n	water velocity at flows below the design flow [ms^{-1}]
W_{grp}	weight of GRP penstock [kg]
W_{steel}	weight of steel penstock [kg]
ϵ	roughness height [0.029mm for GRP]

1 Introduction

The IPCC Climate Change Synthesis Report 2014 (IPCC, 2014) stated that the period from 1983 to 2012 was *likely* the warmest 30-year period of the last 1400 years in the Northern Hemisphere. With high confidence, the report states that oceanic uptake of atmospheric CO_2 is acidifying the oceans, the Greenland and Antarctic ice sheets have lost significant mass (particularly since 1992 and at an even higher rate since 2002) and glaciers worldwide are shrinking. Global mean sea level has risen almost 0.2m since the beginning of the 20th century and precipitation has increased over mid-latitudes.

According to the report, at least half of the observed increase in global average surface temperature since 1951 is *extremely likely* due to anthropogenic emissions of greenhouse gases (including carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O)) and other anthropogenic forcings. At least 80% of CO_2 equivalent emissions since 1970 are due to fossil fuel combustion and industrial processes, with increases driven primarily by global growth in the economy and population. Climate change impacts most strongly on natural systems, for example by altering hydrological systems, shifting the geographic range of species and affecting human society through reduction of crop yields. It is likely that global warming also leads to an increase in extreme events such as heat waves, cyclones, droughts, floods and wildfires.

Although temperatures are expected to continue to rise for some time to come, one major mitigation strategy is to decarbonise primary energy supply and electricity generation, looking towards increased use of renewables, nuclear power and carbon capture and storage (CCS) technologies. Renewable energy supply formed a 10% share of consumed global primary energy in 2016, of which nearly 70% was hydroelectric (i.e. 7% of global primary energy is generated by hydroelectric plants) (BP, 2017a). However, mitigation strategies to maintain CO_2 levels would require low carbon energy supply to be in the region of 20% by 2030, between 40% and 60% by 2050 and between 75% and 100% by 2100 (IPCC, 2014).

Hydropower contributed approximately 16% of electricity generation during 2014 (World Bank, 2017) but estimates of untapped resource are in the region of 10,000TWh per year i.e.

approximately half of current total global electricity generation (World Energy Council, 2017).

China is an interesting case. With its rapidly increasing population, a GDP that has been growing at an exponential rate and an increasing demand for energy, the country would benefit enormously from further exploitation of its water resources..

Currently, almost 9% of primary energy generation in China is by hydroelectric plants and the country hosts over a quarter of the world's hydroelectric supply. China's 13th Renewable Energy Development Five Year Plan (2016 to 2020) (NEA, 2016) aims to increase non-fossil energy supply to 15% by 2020 and to 20% by 2030, increase installed capacity of renewable energy generation to 680GW by 2020 and decrease reliance on foreign companies by continued development of China's own renewable energy industry.

The mountainous terrain and long flowing rivers of China, particularly in the south-eastern sector around the Yangtze basin, makes for rich hydropower resource. Although many large power stations have already been built there and are currently in use, including the world's largest power station, the 22GW Three Gorges Dam, there is still much unexploited resource, particularly if considering 'run-of-river' hydropower. 'Run-of-river' plants provide hydroelectric power with no, or limited, water storage and therefore are subject to seasonal river flows. This hydropower plant type is generally considered more environmentally benign than large dam projects (see section 2.7).

The aim of this study is to provide a methodology to assess the hydropower resource of large river catchments, using the Yangtze drainage basin as a test case, which could also be used to assess the hydropower resource of any global catchment. Hydropower is site specific and each installation is unique, as are the associated costs and feasibility. Local geography and changing hydrology over many seasons are important factors in the likelihood of an installation being economically feasible, and hence high-resolution input datasets are necessary within the constraints of availability and computing power.

This project aims to develop two main algorithms with a view to searching for economically viable hydropower resource across the Yangtze drainage basin:

1) A gridded hydrological model to estimate daily river flows across a catchment, incorporating long time-series of meteorological data and a digital elevation model of the

catchment. This will primarily be required to generate flow duration curves at each grid cell classed as river.

2) A hydropower search algorithm incorporating outputs from the hydrological model, i.e. flow duration curves. The hydropower search algorithm requires the ability to trial many different scheme configurations, such as different turbine types and penstock arrangements, to find an optimum configuration which can be tailored to the user's requirements (e.g. generating maximum profit or maximum energy).

Ultimately these models will be applied to the Yangtze drainage basin with the aim of learning lessons for future development. A further aim of the project is to produce a model that can operate with datasets available within the public domain. This would enable a wider community to utilise the model, whether governments, businesses, academia or the public.

1.1 Research objectives

The hypothesis of this project is that an effective search tool to find economical run-of-river hydropower resource across the Yangtze drainage basin can be developed utilising data sources found within the public domain. As a research question this can be rephrased as:

How can we develop a model to search for economical run-of-river hydropower resource across large river catchments, using the Yangtze drainage basin as a test case?

The key aims of this project are therefore to simulate the flow variation of rivers across the Yangtze basin (known as flow duration curves) by developing a hydrological model, and to incorporate this data within a geospatial hydropower search and costing model.

The objectives of this work therefore include:

- Select or develop the necessary meteorological data inputs for a gridded hydrological model.
- Determine an appropriate snowmelt model.
- Develop a gridded hydrological model to determine surface runoff (river flow) variations across the Yangtze drainage basin and calibrate to obtain a good fit between modelled and observed flows.
- Generate flow duration curve data necessary for the hydropower search algorithm.
- Develop a hydropower search algorithm to determine the economic performance of 'virtual hydropower schemes' across the river network

- Develop a hydropower conflict algorithm to select the most profitable sites and remove sites that conflict spatially to identify the best network of hydropower stations within the catchment.

1.2 Contribution to knowledge

As primary energy consumption is rising, and the increasing electrification of society, there is a great need for an environmentally friendly, clean and abundant electricity supply. Run-of-river hydropower resource, particularly on small rivers, is relatively environmentally benign and can be installed to power local communities or as part of the wider electricity network. A model has been developed within this study which can search for viable hydropower locations across the world's largest catchments (e.g. the Yangtze), which could also operate on catchments of much smaller river basins as there is no lower limit to the size of the catchment (the speed of the model increases as catchment size decreases).

With the key input datasets available within the public domain, or at least developed from sources publicly available, this makes the model appeal to a wide community, not just those who have access to closely guarded, or expensive, datasets.

A distributed hydrological model, estimating river flows in high-resolution spatially and temporally, is a powerful resource and provides the necessary input data to calculate run-of-river hydropower energy generation and associated costs. As part of this model development, a daily evapotranspiration dataset was created, something previously unavailable within the public domain for the Yangtze basin. Through development in the R programming language, the code is both powerful but easy to customise for different catchments and different input data.

A hydropower search methodology that can iterate different designs to optimise a scheme is a powerful tool, particularly as governments, businesses and communities try to make the most of their energy resource. Adopting the RETScreen costing methodology enables potential schemes to be assessed, with predicted costs comparable to those of actual real-world schemes published elsewhere, giving confidence in the model. Maps of viable hydropower locations within the Yangtze are presented within this thesis, and a full list of viable schemes presented on the accompanying DVD.

1.3 Thesis outline

There are 8 chapters within this thesis and an accompanying DVD. This introduction chapter provides the scope and objectives of the work.

Chapter 2 gives an insight into trends and future projections in primary energy consumption and generation, globally and in China. The chapter discusses how fossil-fuel based energy generation produces carbon dioxide emissions which are linked to climate change, therefore supporting the case for the increased role of renewables. Various aspects of hydropower, hydropower resource potential and current exploitation are introduced. A review of hydropower resource assessment is presented alongside current costs and costing methodologies, and an introduction to the characteristics of the Yangtze basin. An overview of hydrology and hydrological modelling is presented together with a brief overview of the methodology used in this project.

Chapter 3 defines the key requirements for development of a hydropower search algorithm, and explains the selection and/or development of input datasets necessary for the search algorithm. These include elevation data and a digital elevation model representation of the Yangtze drainage basin, a lake and reservoir dataset, an introduction to how to determine river flow variations across all rivers within the catchment (i.e. flow duration curves), river width and depth determination, a road dataset and a representation of the electricity transmission network. Finally, a power production and costing methodology is selected.

Chapter 4 details the development of necessary input meteorological datasets used in the hydrological modelling process. This includes selection, and preparation, of daily precipitation and temperature datasets, and development of daily evapotranspiration datasets by combining other meteorological data. A snowmelt model is also selected and discussed.

Chapter 5 details the development of the hydrological model, beginning by introducing hydrological modelling and calibration processes in general. An appropriate hydrological model is selected, G2G, and the structure and description of a new code developed in the R programming language is introduced, integrating the published G2G equations and the input datasets developed in chapters 3 and 4. Calibration is necessary due to the many unknown parameters within the model, and a process is outlined to find an optimum parameter set for the whole of the Yangtze drainage basin via calibration of several smaller sub-catchments.

Flow duration curve data is extracted from nearly 30 years of daily flow data from across the entire Yangtze catchment.

Chapter 6 introduces the RETScreen power estimation and costing methodology, which is integrated into a costing model developed in the R programming language, together with necessary input datasets generated in chapters 3 and from the output of chapter 5. A description of the code and project economics is presented. A full iterative search trialling many 'virtual hydropower stations' is carried out across the Yangtze basin, and a conflict algorithm developed to optimise the hydropower network.

Chapter 7 presents the results of the search, and the characteristics of viable configurations. The accuracy of the estimated gross head is tested, and an alternative method developed due to over-estimation of gross head. An analysis of the costs is presented and compared to other studies within the literature, together with a cost sensitivity analysis of project variables. Power time-series are presented from various example schemes and certain characteristics explored further including the access road distance and transmission costing.

Chapter 8 presents a summary of the findings of the project, including an evaluation of the model and its limitations, as well as discussions for further work to be carried out.

The accompanying DVD includes a full set of results of viable hydropower schemes and the model codes.

2 Literature survey and project background

2.1 Introduction

This chapter aims to review the existing literature and set the context and background of the project. To understand the need for economical hydropower resource mapping, it is necessary to understand primary energy consumption trends. The scientific literature supports the hypothesis that climate change is inextricably linked to primary energy consumption. Therefore, this chapter aims to give insight into the drivers and impacts of climate change and a brief review on the potential climate-related impact on hydropower. An introduction to hydropower will be presented together with current exploitation of hydropower across the globe and in China, and an estimation of future growth. This work builds on previous studies of hydropower resource elsewhere and hence it is important to review the literature of hydropower assessment and costing methodologies. Large dam projects are controversial due to their impact both on ecosystems and human society, whereas small hydro is generally seen as relatively environmentally benign and therefore a brief overview of environmental impacts is important background material. A geographical overview of the Yangtze drainage basin and precipitation in the region gives an insight into the hydropower resource available. Hydrology, meteorology and the use of geographical information systems are all necessary elements of assessing hydropower resource and will be discussed briefly. Finally, this chapter will give an overview of the methodology of the project.

2.2 Global primary energy consumption

Global primary energy use continues to rise, with the world consuming 13,276 million tonnes of oil equivalent (Mtoe) in 2016 (BP, 2017a), equivalent to approximately 154 petawatt hours (PWh) in one year ($1PWh = 1*10^{15}Wh$). Despite a focus on energy efficiency improvements, both to reduce costs and combat climate change (see section 2.3), primary energy consumption has continued to rise over the last decade at a mean growth rate of 1.8% per annum and will

likely continue to do so over the next 20 years, albeit at a slowing growth rate (BP, 2017b) (see figure 2.1).

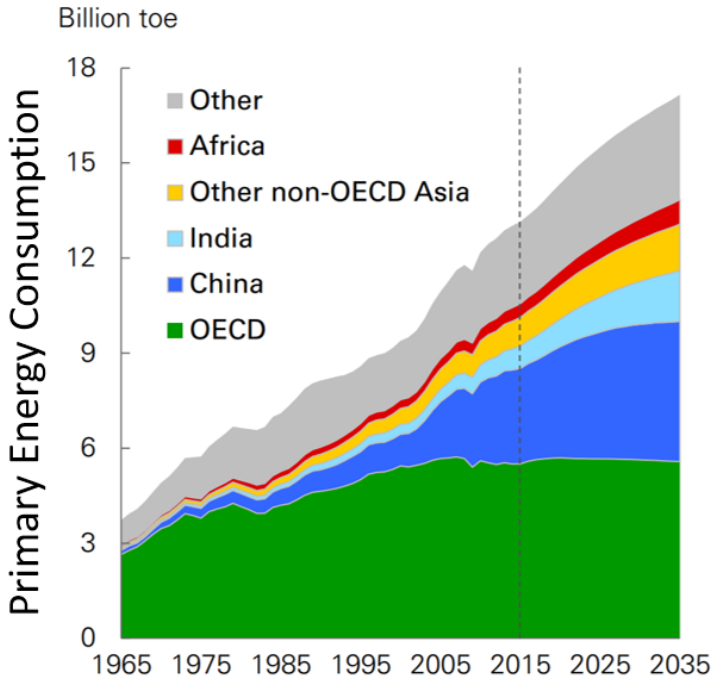


Figure 2.1: Growth in global primary energy consumption by region (BP, 2017b) [Billion tonnes of oil equivalent]

Drivers for primary energy consumption growth are primarily centred on population increase and an increase in global GDP per capita (see figure 2.2). An estimated 1.5 billion extra people are expected to add to the current population by 2035 (an estimated 8.8 billion total). Global GDP is expected to grow at approximately 3% p.a., despite the dramatic slowing in 2008/2009, and the global economy is expected to almost double in 20 years (BP, 2017b; World Bank, 2017). The increase in GDP will account for approximately 75% of primary energy consumption growth as millions are lifted out of poverty from low-income countries. In fact, almost all the growth in world energy demand is from fast-growing emerging economies, with China and India accounting for half of the growth.

Fossil fuel combustion still dominates primary energy generation, forming over 85% of the share (BP, 2017a). Despite an increasing global role for nuclear, hydropower and renewables (which includes wind, solar, geothermal, biomass and biofuels), fossil fuel use will still dominate over the coming years (BP, 2017b) (see figure 2.3).

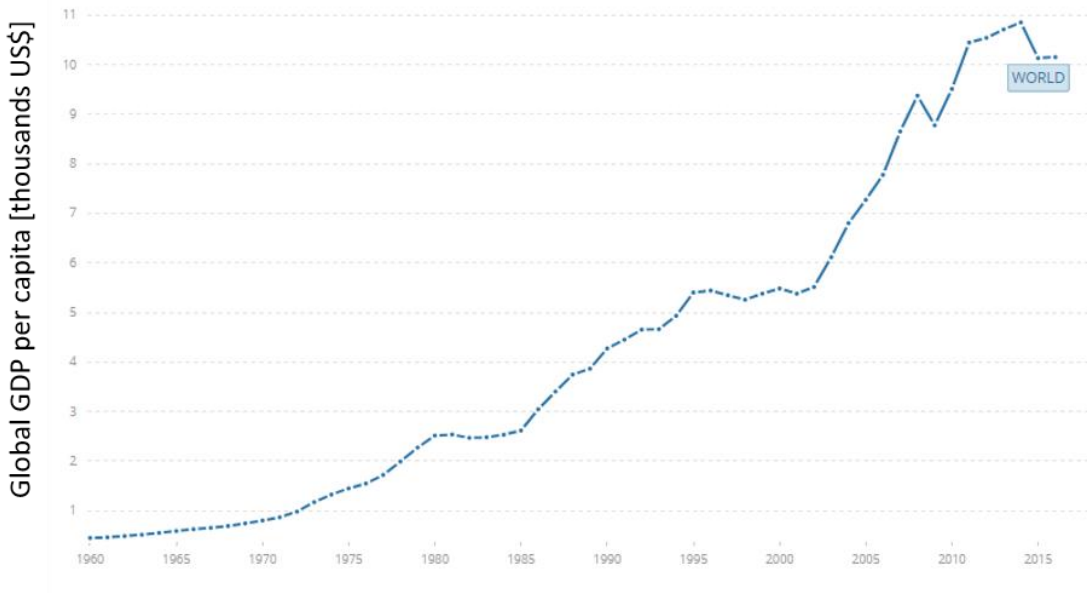


Figure 2.2: Growth in global GDP per capita [thousand US$] (World Bank, 2017)

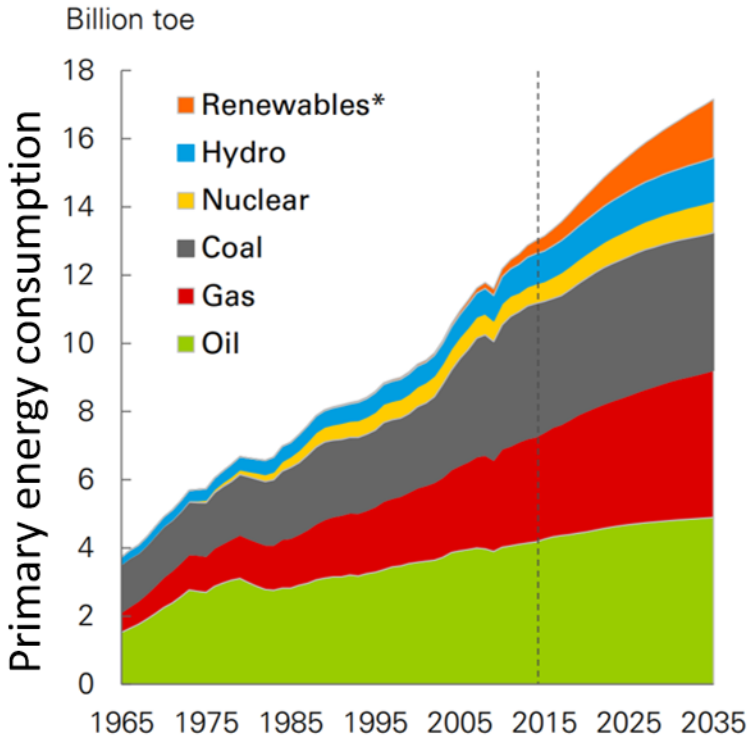


Figure 2.3: Growth in global primary energy consumption by fuel type (BP, 2017b) [Billion tonnes of oil equivalent (toe)]

2.2.1 Primary energy consumption in China

Population growth, and particularly the rise of GDP per capita, are major drivers of primary energy consumption growth in China. The country's GDP growth has been consistently over 6% per annum since the early 1990s (World Bank, 2017) and GDP per capita has increased exponentially (see figure 2.4). Since 1960, the population has nearly doubled and energy use (kg of oil per capita) has tripled since 1990. Electric energy consumption per capita per annum has increased by almost a factor of 8 (see figure 2.5), and is expected to continue to rise, almost reaching parity with OECD countries by 2035, which has levelled out at approximately 9MWh per capita per year (BP, 2017b) (see figure 2.6).

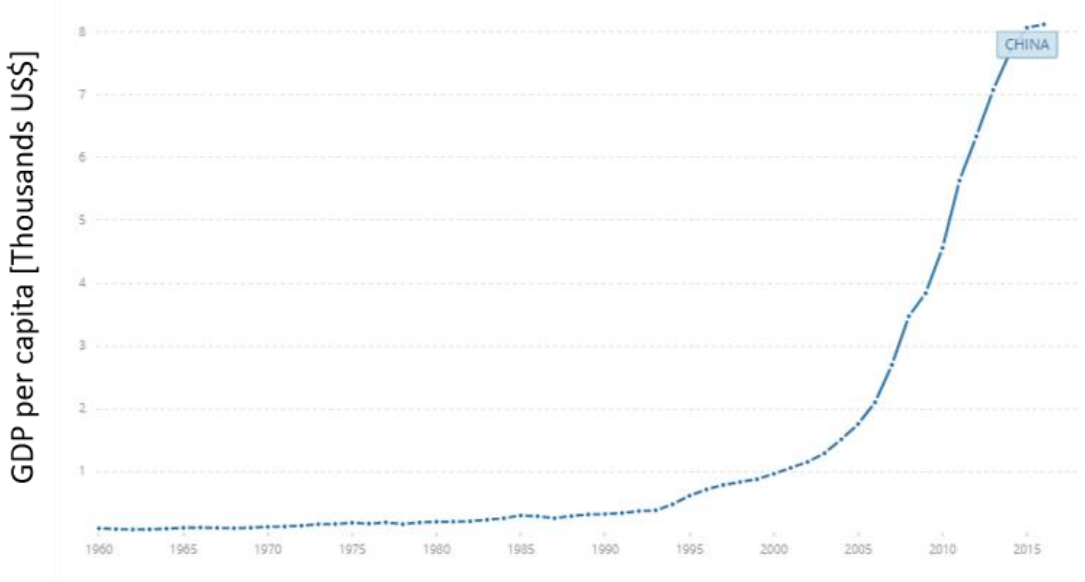


Figure 2.4: Growth of GDP per capita in China [thousands of US$] (World Bank, 2017)

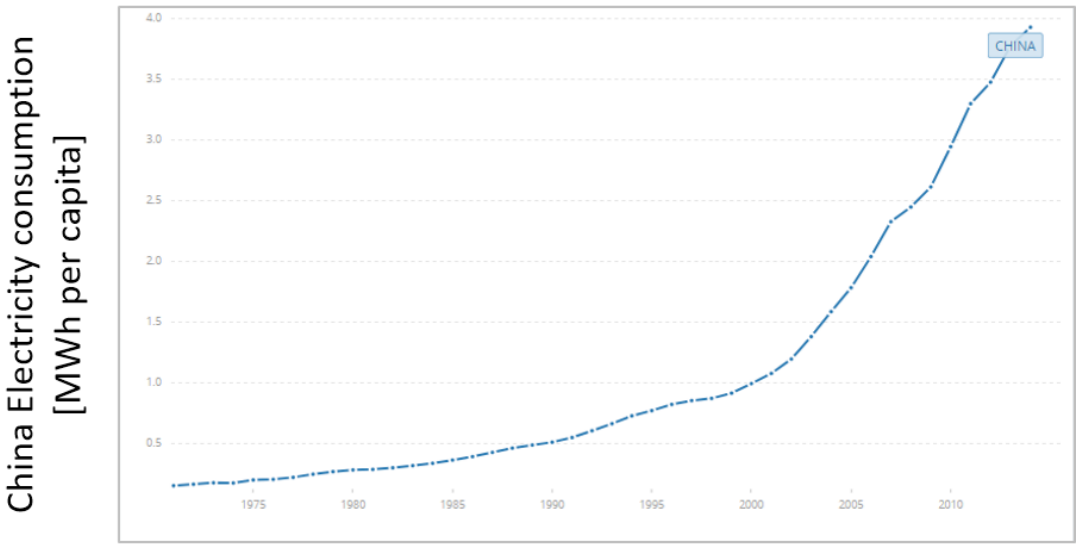


Figure 2.5: Growth in electrical energy consumption in China [MWh per capita] (World Bank, 2017)

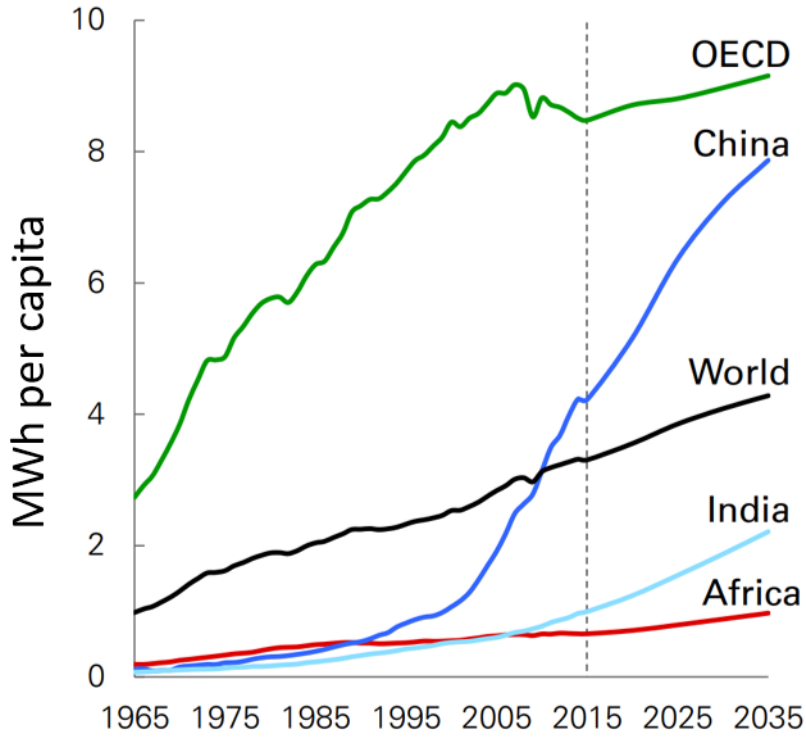


Figure 2.6: Regional electricity consumption [MWh per capita per annum] and outlook (BP, 2017b)

Unsurprisingly, China's energy use is dominated by fossil fuels, particularly coal (see figure 2.7). Mean ambient air quality (measured as micrograms per cubic metre PM2.5) is approximately 5 times higher than in the UK (Brauer et al., 2015), a significant amount (approximately 40%) of which is due to coal combustion (Ma et al., 2016; Quan et al., 2014). However, China's share of renewables and hydroelectric power is above the global mean, and it is estimated that China's renewable power is the largest source of global growth in renewables, expected to add more over the next 20 years than the EU and US combined (BP, 2017b). Furthermore, China's energy intensity, or energy use per $ GDP, has decreased by a factor of almost 3 between 1990 and 2016.

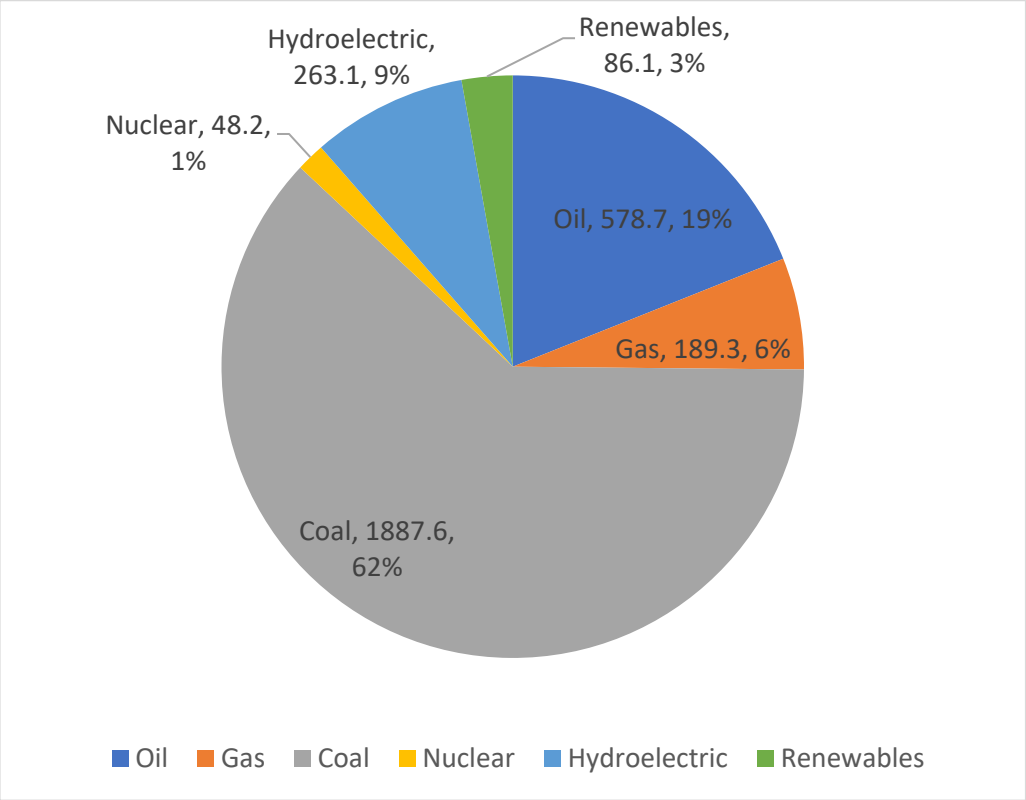


Figure 2.7: Share of China's primary energy source (BP, 2017a) [Million tonnes of oil equivalent (Mtoe)]

2.3 Climate change and the case for renewables

There is an overwhelming body of scientific literature (although not everybody agrees) that shows the climate is warming at an unprecedented rate, attributed to anthropogenic (man-made) emissions of greenhouse gases. The Intergovernmental Panel on Climate Change (IPCC) declared in their synthesis report: *Summary for policy makers* (IPCC, 2014) that human influence on the climate system is clear and recent anthropogenic emissions of greenhouse gases are the highest in history. They state that recent climate changes have had widespread impact on human and natural systems, and this warming of the planet is unequivocal and unprecedented, with most of the severe warming occurring since the 1950s.

Each of the last three decades has been successively warmer at the Earth's surface, and the period from 1983 to 2012 was *likely* the warmest 30-year period of the last 1400 years in the Northern Hemisphere (assessed with *medium confidence*). The global mean temperature has warmed by approximately 0.85°C from 1880 to 2012. Oceans have warmed significantly, particularly in the upper 75m, leading to increased precipitation in mid-latitudes and increased evaporation in some regions (assessed with *high confidence*).

The uptake of CO_2 has decreased the pH of the oceans (i.e. the oceans are acidifying). NASA (2017) estimated that Antarctica has lost approximately 2000 gigatonnes (Gt) of ice-mass since 2002 and Greenland lost nearly 4000 Gt in the same period. Arctic sea-ice is decreasing at a rate of 13% per year loss since 1980, reaching an all-time minimum in 2012. Mountain glaciers across the globe are shrinking (Bliss et al., 2013; Roe et al., 2017). This, together with expansion of the oceans due to warming temperatures, has led to a sea-level rise of approximately 0.2m since 1900.

At least half of this climatic warming, according to the IPCC, is *extremely likely* due to anthropogenic greenhouse gas emissions of carbon dioxide (CO_2), methane (CH_4) and nitrous oxides (N_2O), together with other anthropogenic causes such as emissions of black carbon (Ramanathan and Carmichael, 2008; Bond et al., 2013). Increases in emissions since 1970 are shown in figure 2.8.

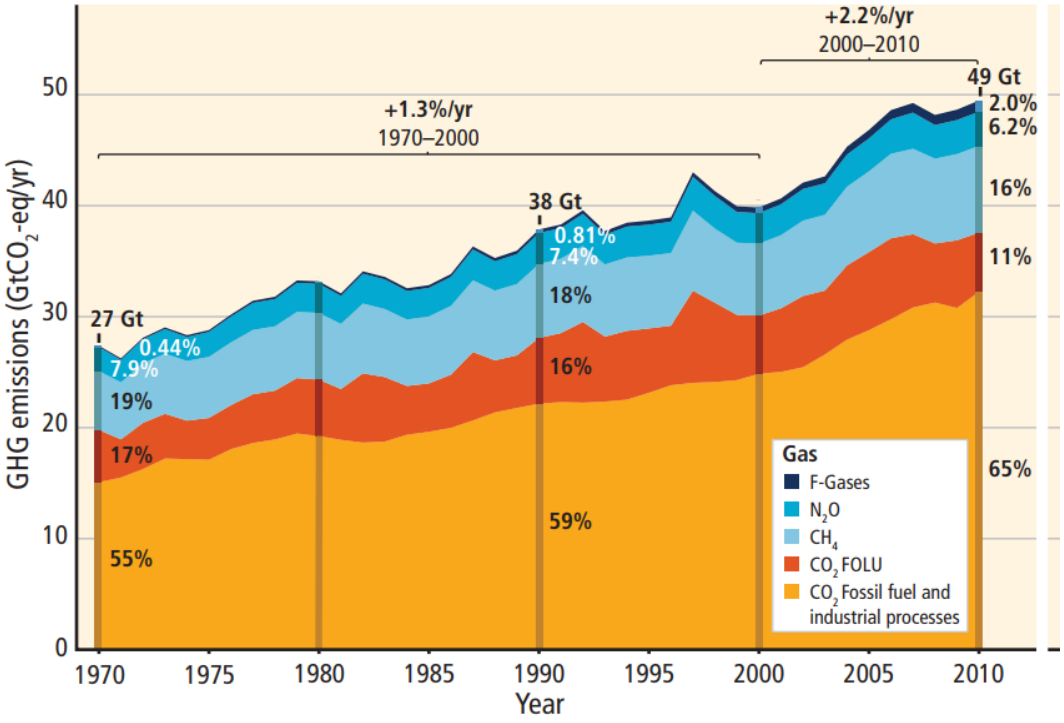


Figure 2.8: Anthropogenic greenhouse gas emissions (GHG) between 1970 and 2010 (CO_2 FOLU represents CO_2 emissions from land and forestry use; F gases represents Fluorinated gases such as chlorofluorocarbons, hydrochlorofluorocarbons and hydrofluorocarbons (IPCC, 2014))

Climate change impacts most strongly on natural systems, particularly hydrological systems affecting water resources and terrestrial, freshwater and marine species. Human crop yields have been affected (negatively) and there is *very likely* an increase in extreme events due to climate change (e.g. heat waves, cyclones, droughts, floods and wildfires).

Key drivers of climate change include population increase, increased economic activity, lifestyle changes, increased energy use, new technology and climate policy – a similar set of drivers to those driving increases in primary energy. Hence, there is a need for increased clean energy generation, either through renewable energy generation, nuclear power or carbon capture of CO_2 from fossil-fuel emissions.

Any impact on the climatic system will impact on the water cycle (Kundzewicz, 2008), Therefore there may be am impact on hydropower installations, modifying the performance of existing schemes or the expected performance of schemes in the future. There are many studies on climate change and the impact on water resources and/or hydropower. Hamududu and Killingtveit (2012) assessed the impact on global hydropower and concluded that very little impact will occur on global generation, but some sites will see decreased annual runoff

and others will see a greater amount. Agrawala et al. (2003) concluded that hydropower in Nepal, which relies heavily on hydropower for electricity generation (91%), will encounter higher storm flows possibly causing damage to hydro equipment, and lower dry season flows due to glacial retreat.

In a study of Canadian water resources (Minville et al., 2009) employed a coupled regional climate model with the SRES A2 (IPCC Special Report on Emissions) emissions scenario and found that hydropower output would decrease by 2% up until 2039 but then increase by 9% after 2040. A study on climate change and hydropower production in the Swiss Alps (Schaefli et al., 2007) concluded a statistically significant negative impact on current system performance. China's hydropower generation is also vulnerable to climate change, potentially losing between 2% to 3% by 2020 and 4% to 6% by 2030 unless mitigation action is taken (Wang et al., 2014). However, some regions would be hit harder than others, with Sichuan, Yunnan and Gansu provinces (which fall at least in part within the Yangtze watershed) most adversely affected.

2.4 Characteristics of hydropower

Hydropower is the conversion of energy (potential or kinetic) stored within falling, or fast-moving, water to provide mechanical or electrical power. In modern terms, hydropower is thought of almost exclusively as hydroelectric power although some mechanical systems still exist. High efficiency electricity generating turbines have largely replaced the water wheel, with the power available determined by:

$$P = \rho g h Q \eta \qquad (2.1)$$

where P is power [W], ρ is the density of water [kgm^{-3}], g is the acceleration due to gravity [ms^{-2}], h is the head [m], Q the flow through the turbine [m^3s^{-1}] and η is the efficiency of the system. A typical small scheme is shown in figure 2.9, with an intake diverting water to a penstock pipe, channelling water into a turbine linked to a generator (within a powerhouse) to convert the mechanical energy to electrical, which is then transmitted to the electricity grid or to a local electricity network.

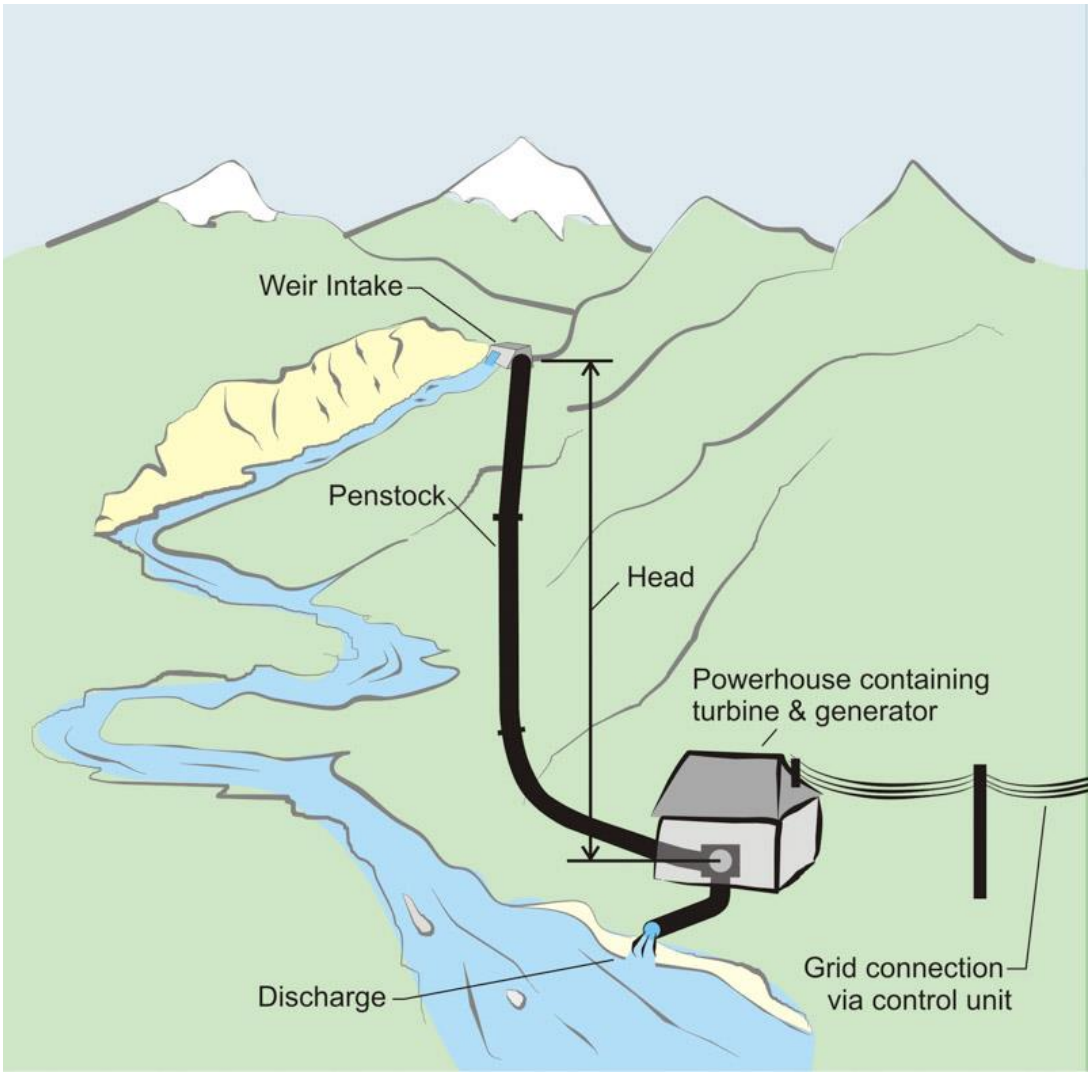


Figure 2.9: Layout of a typical small hydropower scheme (Snowdonia Hydro, 2014)

Losses occur due to friction within the penstock (known as head-loss due to reducing the effective head), and at the generator, transformer and through transmission. Hydraulic turbine efficiencies are often in the range of 80% to over 90% (Paish, 2002). It is important to note that these efficiencies are only achievable under optimal conditions, and depend on flow through the turbine and the output demand (Boyle, 2004). Turbines are selected with a 'design flow' (i.e. maximum flow) in mind, with efficiency of the turbine peaking at flows just below the design flow. Hydropower schemes with a reservoir can, to some extent, control the flow through the turbines to maximise turbine efficiency, but not so in run-of-river schemes which are subject to the variable flow of rivers (see 'Run-of-river hydropower' section 2.4.1).

In relatively simple assessments of hydropower resource, the mean annual (or multi-annual) flow is used together with the head (difference in height between the intake of the scheme and turbine outlet) in equation 2.1 to calculate the potential power. However, with variable flow schemes the calculation is far more complex.

Hydropower plant installed capacities (the maximum power output of a plant) range from a few kW (in pico-hydro schemes) to several GW. The Three Gorges Dam on the Yangtze River, China, is the largest power station in the world at 22GW installed capacity due to its high dam and the huge volume of flowing water. In this assessment, the terms micro, mini, small, large and very-large hydropower will be used to represent different plant installed capacity categories (where micro is less than 100kW, mini is between 100kW to 1MW, small is between 1MW to 50MW, large is between 50MW and 1GW and very-large is greater than 1GW). Note that the term 'small' can have different meanings in different countries. For example, only the largest of Scotland's hydropower schemes have a capacity greater than 50MW but in China this would be classed as small.

Hydropower stations are usually classed as low, medium or high head – where head represents the height difference between where the water enters into a hydropower system and where it leaves. High head sites are usually installed in mountain (or hill) regions and have a head of 100m or more, with water diverted into a penstock (a pipe), channelling water into a turbine, before being returned to the river. The term 'penstock' originated due to the historical method of using a gate, or stock, to control penned up water. Medium head sites may include a high dam to hold a reservoir, and therefore add head to the scheme, with the penstock channelled through the dam wall to the river outlet at the base. Low head sites utilise a low dam or barrage, relying on the large river flows to generate power rather than head.

2.4.1 Run-of-river hydropower

Run-of-river (ROR) hydropower is where little or no water storage is provided, hence the amount of water flowing through the scheme depends upon the river flow at any one time. Some schemes employ limited storage to allow buffering of variability, preserving efficiency due to allowing schemes to operate for longer at design flow. Rivers vary greatly in their flow, reaching a peak following storm events and snowmelt in spring, and experience a decline during drought conditions. Hence ROR hydropower is an intermittent energy resource, and the amount generated depends upon the availability of flow and the design of the scheme.

A configuration with a turbine designed to perform at high efficiency under low flow conditions may generate a constant supply of energy for much of the year, but only generate a low output. Schemes with turbines designed for high flow conditions i.e. flood flows, may generate a large amount of energy but only for very short periods of the year, and would be more expensive due to the higher costs of larger components. When the river returns to low flow again, the scheme would generate little or no power due to low efficiency of the turbine. Hence there is usually a cost/energy trade off, with schemes usually designed in between the two extremes.

A low dam, or weir, is usually provided to ensure sufficient water enters the penstock, although a minimum amount is left in the river for environmental considerations. ROR schemes are generally considered environmentally benign compared to large hydropower projects with a large reservoir, but do have some environmental impacts (see section 2.7). They are also usually of lower cost compared to large hydro dam projects. As flow varies both seasonally and annually, to fully cost potential ROR projects requires long time-series of flow data to generate a flow duration curve (FDC). FDCs are usually generated over a minimum of 25-30 years to account for climatic variation. Furthermore, many of the likely locations for ROR projects, particularly in remote regions, will be ungauged (i.e. the flow of the river not measured by a meter), requiring other methods to estimate the flow regime.

Only ROR-type projects will be considered in this thesis. Also, the installed dam will not create extra head and the net head will only be due to the natural geography between the intake and powerhouse. However, many real-world projects classed as ROR may have a significant dam. For example, the 2620MW Chief Joseph Dam on the Columbia River (Washington State, USA) is classed as ROR and has a 72m-high dam, creating a significant head.

2.4.2 Turbine types

Water turbines are usually separated into impulse and reaction types. There are wide design variations within each type to meet the specific hydraulic conditions of a scheme, with most modern turbines designed to generate electricity as opposed to mechanical power.

Impulse turbines convert the potential energy of water to kinetic energy by discharging water into the air through carefully shaped nozzles, directed onto curved buckets fixed on the rotating *runner* to extract useful work. Most impulse type turbines are based on the Pelton wheel design where the free water jet strikes buckets tangentially. Each bucket has a high

ridge, or splitter, to divide the incoming jet onto the two halves of the bucket. The Pelton wheel can capture virtually all the energy of the incoming water due to almost 180° rotation of the water stream. Multiple jets can be added, particularly when the turbine is mounted vertically, leading to high efficiency. To adjust quickly to changes in electrical load, the water flow rate is controlled by a spear or needle valve that slides forward or backward by a hydraulic servomotor. Pelton turbines are designed for high head schemes (typically between 80m and 1000m) with relatively low water flows.

In Turgo turbines, the jet impinges at an oblique angle on the runner from one side and discharges at the other, also used in moderately high heads (typically between 50m and 250m). Another impulse type is the Cross-flow turbine, where the water passes transversely, or across, the turbine blades, suited for low heads (typically between 5m and 100m) and high flows. Due to their flat efficiency curve and low price they are popular in small ROR hydropower schemes.

In reaction turbines, the turbine blades are totally submersed in the flow of the water and are enclosed within a pressurised casing. A reaction turbine is powered mainly by the change in pressure, called a "pressure drop" across the casings body as this reduction in water pressure and velocity releases energy causing a reaction (hence the name) by moving the turbines blades. Although unusual, the flow of water through a reaction turbine may be reversed due to the angle of the internal blades, so a reaction turbine can also be used to pump water and vice versa (Alternative Energy Tutorials, 2018). However, this could result in poor efficiency in smaller schemes, and it is more normal to have dual pump turbines each optimised to drive or be driven (Wallace, 2018).

A propeller turbine has a large propeller turned by the flow of the water extracting energy. Kaplan turbines are like propeller turbines but with adjustable blades to achieve efficiency over a wide range of flow and heads. Both propeller and Kaplan types are designed for low heads (usually below 25m), and have spiral-shaped inlet tubes with inlet guide vanes that wrap around the wicket gate. Water is directed through the wicket gate and spirals onto a propeller-shaped runner.

The most common turbine in operation is the Francis turbine due to its use in a wide range of heads (from 12m to 600m) and flows. Most commonly the Francis turbine is designed with a welded or cast steel spiral casing distributing water evenly to all the inlet gates which can be fully opened or closed depending on the desired power output. Openings in the spiral casing

convert the pressure energy of the fluid into momentum before impinging on the blades. In all the reaction turbine designs there is a draft tube connecting the runner exit to the tail race (the flume or channel leading water away from the turbine back to the river) to reduce the velocity of the discharged water and thus minimise kinetic energy losses.

2.4.3 Hydropower electricity generation globally and in China

Hydropower production formed 7% and 9% respectively of global and China primary energy generation in 2016. In terms of global hydropower production, China is a world leader, generating 263 Mtoe (approximately 3 PWhr) in 2016 (BP, 2017a), eclipsing not only every other country but every continent also (see figure 2.10).

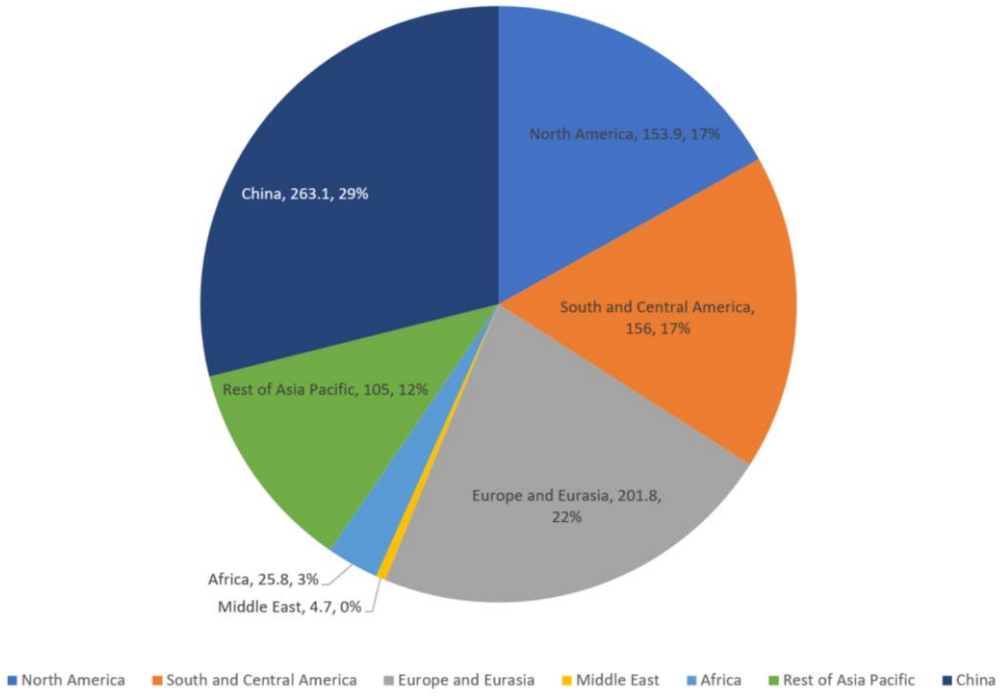


Figure 2.10: Global hydropower production by region [Mtoe] (BP, 2017a)

World installed hydropower capacity in 2016 reached 1,245GW, generating 16% of global electricity production (IHA, 2017; World Energy Council, 2017). China's installed capacity by 2017 was approximately 331GW, following strong growth in the sector since 2005. The expected additional hydropower resource in the next decade (2015-2025) (globally) is expected to match the additional hydropower implemented between 2005-2015, but not as influenced by China than in the preceding decade (BP, 2017b) (see figure 2.11). Small

hydropower installed capacity is currently estimated to be 78GW globally (UNIDO, 2016), 51% of which is in China (in the UNIDO report 'small' was defined as less than 10MW). Other estimates of China's small hydropower installed capacity were between 60GW and 70GW by 2015 with more than 45,000 stations. (Kong et al., 2015; Cheng et al., 2014) (in these studies small was defined as less than 50MW).

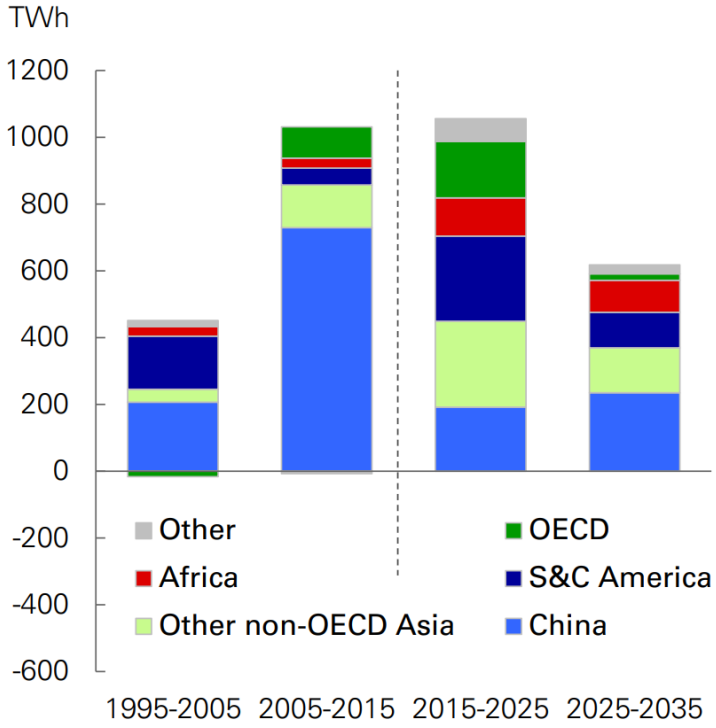


Figure 2.11: Additional hydropower resource per decade by region [TWh] (BP, 2017b)

2.5 Assessments of hydropower resource

Computer-based geospatial processing tools coupled with the availability of high-resolution geospatial datasets has enabled automated hydropower assessment, either at a focused site or across a wide area. Previously this could only be achieved through detailed analysis of maps, a labour-intensive process.

Francfort et al. (1993) developed the 'Hydropower Evaluation Software', as part of the U.S. Department of Energy's assessment of undeveloped hydropower resource in the United States. This was a menu-driven software application allowing a user to assign environmental attributes to potential hydropower sites to produce suitability reports. The final report (Conner et al., 1998) estimated 30GW of economically viable undeveloped hydropower across 5,677

sites. Balance et al. (2000) conducted a preliminary assessment of hydropower resource potential in South Africa from digital maps of slope and runoff, searching for both ROR micro-hydro and a damming model for larger hydropower sites. The approach looked for steep gradients across a 400m grid using a geographical information system (GIS), with flow calculated as the (previously measured) mean annual runoff for each catchment. Although a relatively simple methodology, this enabled areas to be identified for further investigation.

Dudhani et al. (2006) used remote sensing data from the IRS-1D LISS III satellite to map water resources across North India and extract geographical features around sites to give an opinion on its suitability for small-scale hydropower. Hall (2007) utilised GIS tools and analytical modelling of natural streams to perform virtual 'river inventories' to assess the gross power potential of hydropower across the United States. The results of this study fed into a new tool called the 'Virtual Hydropower Prospector'. Annual mean power was estimated by GIS extraction of the gross head of each stream reach and flow of ungauged catchments estimated by regression equations. Economically feasible plants had to fulfil the criteria of being within 1 mile of an access road and within 1 mile of an electrical load, and generate power greater than 10kW.

The Rapid Hydropower Assessment Model (RHAM) (KWL, 2007) assessed the ROR hydroelectric resource for the Province of British Columbia, Canada. Incorporating a digital elevation model (DEM) and mean annual runoff within a unique GIS algorithm, they identified every significant stream and river reach within a given area, their respective flow rates throughout the year and the maximum elevation drop along each reach. In terms of scale, the 950,000km^2 of this study is one of the largest areas assessed in this review. Crépon (2009) re-assessed French hydropower potential at a regional scale in France using an innovative numerical methodology combining GIS, hydrologic and hydrographic characteristics of sub-basins and rainfall maps. The study also considered the potential of utilising non-hydroelectric existing dams.

Hydrobot (Forrest, 2006; 2010) combines GIS and financial assessment tools to estimate flow duration curves at any point of the river network of Scotland, and estimates construction costs and financial benefits. The model was used in a study by the Forum for Renewable Energy Development on behalf of the Scottish Government to assess the nations hydropower potential (Nick Forrest Associates, 2008). Alterach et al. (2009), evaluated the potential hydropower resource in Italy with a numerical technique coupled with GIS, applying hydrological and

hydrographical characteristics of interconnected sub-basins and rainfall maps. Mean rainfall maps processed by the 'spline' interpolation method were used to determine flow using runoff coefficients in conjunction with geodetic heads to estimate power.

Kusre et al. (2010) assessed the hydropower potential of the 2228km^2 Kopili River basin in Assam, India, using GIS and the SWAT (Soil and Water Assessment Tool) hydrological model, integrating a digital elevation model (DEM), stream network, soil map, land use and climate data to determine head and discharge. Flow duration curves were estimated from 10 years of data. Larentis et al. (2010) developed *Hydrospot*, a series of FORTRAN routines integrated with ArcGIS to survey a digitised river map and trial different dam-powerhouse layouts in Brazil. Virtual dams were fitted across rivers according to the local geography and the inundated area of the reservoir behind automatically calculated.

A review of tools available for small hydropower plant resource planning and development was conducted by Punys et al. (2011), stating how the DEM was an especially important step in GIS history, enabling access to the hydrological and morphometric characteristics of a river basin. Feizizadeh and Haslauer (2012) used interpolated precipitation and evapotranspiration data by the Inverse Distance Weighted method to estimate runoff across each grid cell of the Tabriz basin 20m DEM in Iran, and used cell slope to calculate the potential energy potential across the region. Meijer's (2012) Master's thesis investigated global hydropower potential, by estimating runoff at each 3-arcsecond HyroSHEDS DEM grid cell by accumulating GRDC runoff fields annual mean data. The slope across each cell was calculated to estimate the head. Total theoretical hydropower capacity was estimated at 20TW.

Þórarinsdóttir (2012) developed a methodology for estimation of hydropower in Iceland using high-resolution hydrological modelling, employing the WaSiM model (a grid-based Water Flow and Balance Simulation model) to estimate discharge based on gridded precipitation, and elevation data from ArcGIS. The study estimated a FDC for each grid cell and then tested the hydropower potential of each cell at different percentile flows. Kao, S.C. et al. (2014) evaluated the new stream-reach development resource potential of more than 3 million US streams by using flow and head data from numerous sources and discretising the river network into 150m reaches. Hydropower potential was calculated assuming the design flow was Q_{30} (i.e. 70th percentile flow) and flow and height data interpolated for ungauged reaches.

Duncan (2014) investigated Scotland's hydropower resource as a PhD project by gridded hydrological modelling of Scotland's water resources based on the Centre of Ecology and

Hydrology's G2G model. FDC data were generated by analysing daily flow grids at 200m resolution and used as input to a hydropower search methodology. A stream network was developed using ArcHydro (see section 2.11) and a trial intake set at the top of each river reach. A virtual penstock was extended downstream to the next river point, with different design flows tested iteratively, with data stored as an SQL database. Costing was calculated based on RETScreen equations and the configuration with the best Net Present Value (NPV) kept and the others rejected. Net Present Value (NPV) is the difference between the present value of cash inflows and the present value of cash outflows, and used in capital budgeting to analyse the profitability of a projected investment or project. The penstock was then extended further downstream and re-costed up to a maximum of 10km length. The intake location was then moved one step down the river reach and the process repeated. Any conflicting configurations were rejected, selecting the site with best NPV.

Zhou et al. (2015) estimated the global gross, technical, economic and exploitable hydropower resource by simulating runoff over all land grids using the Global Water Availability Model at 0.5 degrees resolution. Head was estimated by the difference in height to the next cell. Costing was estimated by empirical cost equations developed by Hall et al. (2003). They estimated an exploitable installed capacity resource of 1.8TW generating 16PWh per annum. Pandey et al. (2015) assessed the hydropower potential of the Mat river basin (Mizoram, India) by estimating monthly discharge with the SWAT model and a high-resolution DEM.

Soulis et al. (2016) considered hydropower potential at historical hydropower sites in poorly gauged areas by embedding a spatially distributed hydrological model in a geo-information system, assuming a constant stream length of 425m. Chelelgo et al. (2016) modelled micro-hydro potential within the River Perkerra catchment (Nepal) by calculating FDC from historical monthly average discharge data at a gauging station and assuming the amount of runoff is proportional to the flow accumulation at a cell.

2.6 Costs of hydropower

2.6.1 Installation costs of hydropower and Levelised Cost of Electricity (LCOE)

When considering costs of hydropower schemes, it is usual to estimate the costs per installed kW (i.e. the construction costs) and the costs per kWhr (i.e. the operating costs). Operating costs are usually given as the Levelised Cost of Electricity (LCOE), the net present value of

the unit-cost of electricity over the lifetime of a generating asset, often assumed to be the average price the asset must receive to break even over its lifetime.

Published installed costs per kW generally decrease as projects increase in size due to economies of scale as demonstrated in table 2.1.

Size of scheme	**Installed costs per kW (US$)**	**Source**
Small (assumed < 10MW)	1300 – 8000	Renewable Energy Technologies: Cost Analysis Series, Hydropower (IRENA, 2012)
Large	1050 – 7650	IRENA, 2012
Small (assumed < 10MW)	2000 - 4000	Renewable Energy Essentials (IEA, 2010)
Medium (10MW to 100MW)	2000 – 3000	IEA, 2010
Medium (100MW to 300MW)	2000 - 3000	IEA, 2010
Large	< 2000	IEA, 2010
Small (<10MW)	1369 – 9400 (median 5281)	Projected Costs of Generating Electricity 2015 Edition (IEA & NEA, 2015)
Large	598 – 8687 (median 2493)	IEA & NEA, 2015

Table 2.1: Installed costs per kW of hydropower

Hydropower is capital intensive but has low operating costs, with LCOE among the lowest of generation types available. However, many of the optimal sites have already been developed and therefore most of the competitive sites are no longer available. The World Energy Council compared global hydropower capital (CAPEX) and operational costs (LCOE) to other generation types, as shown in table 2.2. Note that the range of costs for each generation type are partly attributable to the economies of different countries (e.g. labour costs are significantly lower in India than Japan).

Generation type	CAPEX [US$ per kW]	LCOE [US$ per MWh]
Onshore wind	1080 to 2450	47 to 136
Offshore wind	4290 to 6080	147 to 367
Solar PV	1450 to 2660	79 to 439
Solar thermal (no storage)	3080 to 7670	123 to 490
Biomass and waste	830 to 7700	34 to 210
Geothermal	1080 to 6070	39 to 276
Tidal	6730 to 16,050	263 to 1049
Wave	5480 to 16050	284 to 1058
Hydro	1400 to 4150	19 to 302
Coal	660 to 3700	35 to 172
Gas	760 to 1510	61 to 148
Nuclear	3570 to 6520	91 to 147

Table 2.2: Comparison of CAPEX costs and levelised costs of electricity for different generation types (World Energy Council, 2013)

2.6.2 Costing of hydropower installations

Empirical formulae for rapid costing of small hydropower schemes (<50MW) considering a scheme's head, flow and/or plant capacity have been published by many authors (Matthias 2001; Papantonis, 2001; Singal et al., 2007, 2008, 2010; Aggidis et al., 2010; Zhang et al., 2012). These studies mostly focus on particular components of a scheme (e.g. the electromagnetic equipment), with the Singal et al. studies being the most complete as they determine both electromagnetic components and civil works costs. However, these studies do not consider other project elements such as engineering, development, access roads and transmission.

HydroHELP (2008) is a series of commercial tools that enable pre-feasibility studies of hydropower schemes based on various earlier studies within the literature (Gordon and Penman, 1979; Gordon, 1981; Gordon, 1983; Gordon and Noel, 1986; Gordon, 2003; Gordon, 2008). These form the basis of the RETScreen power production and costing software (RETScreen, 2004) (see section 3.10). Actual installed capacity, power generation and costs of nine Turkish small hydropower plants were compared to RETScreen pre-feasibility estimates and the results given in figure 2.12 (Yuce and Yuce, 2016).

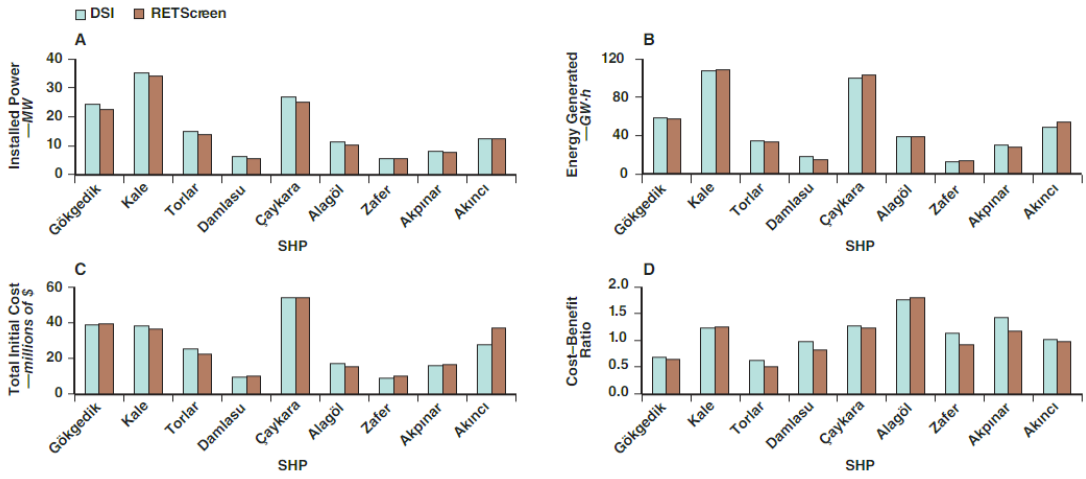


Figure 2.12: Comparison of actual costs of nine small hydropower plants in Turkey given by the State Hydraulic Works of Turkey (DSI) and RETSceen (Yuce and Yuce, 2016)

2.6.3 Financial risk and economic optimisation

There is financial risk in hydropower design due to the potential of reduced surface runoff by overestimating natural meteorological patterns and/or a changing climate, as most schemes are designed based on recent climate history (Alavian et al. 2009; Gjermundsen and Jenssen 2001; Mimikou and Baltas 1997; Harrison and Whitington 2001, 2003; Sample et al., 2015). Milly et al. (2008) concluded that the idea that natural systems fluctuate within an unchanging envelope of variability is no longer valid. Therefore, there is the possibility of over-design with expensive components designed for higher flows or under-design relative to the probability of extreme inflow events (International Rivers, 2012). Assessments rarely evaluate power generation and associated revenue changes associated with climate change. Hydro-dependency may also result in purchase of energy generated by other sources at premium prices when water levels are low.

Kucukali (2011) developed a fuzzy rating tool (a concept which implies gradations of significance or unsharp (variable) boundaries of application) to provide a flexible and easily understood methodology to analyse project risk. Risks considered included site geology, land use, environmental issues, grid connection, social acceptance, macroeconomic reasons, natural hazards, changes of laws and regulations, terrorism, access to infrastructure and revenue. Of these risks, environmental issues and site geology were considered the most important. There is also a risk of catastrophic failure of dams with potential for large costs

both financially and in fatalities (Sovacool et al., 2015). Cost and time overrun is commonplace. In a sample of 61 hydroelectric dam projects, 75% of projects had a mean time overrun of 60% greater than expected with a mean cost overrun of nearly 2.5 billion US$ (Sovacool et al., 2014).

Cost optimisation of projects aims to design the project usually to maximise NPV and/or minimise cost per installed kW and LCOE. Cost optimisation has led to development of various software techniques including Decision Support System (Sharma et al., 2015), Particle Swarm Optimisation (Rahi et al., 2012) and use of the Bat algorithm (Gupta and Sharma, 2016). Optimal sizing of hydropower components has been investigated by Anagnostopoulos and Papantonis (2007) who presented a numerical method to select the design of two turbines operating in parallel which can be of different type and size, concluding that two turbines of differing design sufficiently improves the economics and energy production of a plant. This is due to enabling a plant to perform at or near to design flow to either one of the turbines with the other shut down, or near to design flow of both together. Alexander and Giddens (2008) and Edeoja et al. (2016) investigated penstock sizing to optimise the performance of micro- and pico-hydropower schemes.

2.7 Hydropower and the environment

Due to its provision of low cost and clean electricity, hydropower is an obvious energy choice, but there is still – as with all energy generation choices - some impact on the environment. Therefore, an obvious question is how much hydropower can be exploited without causing widespread environmental damage (WWF, 2004). The United Nations (UN, 2003) reported 60% of the world's largest rivers, including the Yangtze, are already severely fragmented by dams, diversions and canals leading to a degradation of ecosystems. Ecosystem impact is almost always associated with social impact due to degradation of water supply, fisheries, flood control and flood plain fertility – and often it is the poor who are most affected when ecosystems are modified.

Potential negative impacts of both small and large dam projects include (EHSA, 2010):

- Habitat modification due to the interruption of spatial and temporal continuity of rivers
- A barrier to fish passage

- Death of fish due to passing through the turbine
- Visual intrusion
- A barrier to the movement of other animals
- Alteration of natural habitat conditions
- Noise emission

A further impact is the trapping of sediment behind dams (Kuenzer et al., 2013), which leads to areas downstream being deprived of essential nutrients for biodiversity and crop growth. Large dam construction has other impacts not associated with small dams, including (Xu et al., 2013):

- Displacement of people due to reservoir development
- A shortage of human carrying capacity (the maximum population size that the environment can sustain indefinitely) for the displaced people in the area of the dam
- Soil erosion in the reservoir area
- Eutrophication (enrichment of a water body with nutrients which can lead to oxygen depletion due to excessive growth of algae and plants) of the reservoir area
- Downstream riverbed erosion
- Reservoir bank stability
- Impact on lakes downstream
- Methane emissions (a powerful greenhouse gas) due to decomposition of biomass from stratified reservoirs where the bottom layers are anoxic (Lin, 2011)

Xu et al. (2013) note in their study of the environmental impact of the Three Gorges Dam that there is significantly more environmental impact than anticipated in the Environmental Impact Statement of the Yangtze Three Gorges Project (EIADCAS and RIPYWR, 1995). Lin (2011) stated that reservoir creation can increase incidence of disease in humans due to the increase of breeding grounds for disease-carrying organisms, particularly in tropical climates (e.g. mosquitoes, vectors for malaria and snails, vectors for schistosomiasis).

Lin (2011) also claimed that methane emissions following flooding of wide areas of vegetation to form reservoirs in some tropical locations (e.g. Brazil) can emit 3.5 times the CO_2 equivalent of fossil fuel combustion plant of the same generating capacity. However, it is estimated that a large majority of methane emissions could be recovered from large dams and used for power generation (Lima et al., 2008). Methods include installation of a barrier device called a “gate-

buoy" which diverts methane depleted surface waters to the turbine, allowing extraction of methane from deeper waters (particularly the hypolimnion) by bubbling or spraying devices, extracting the methane into a sealed vessel.

ROR hydropower is often labelled as environmentally benign, but a review of ROR environmental impacts include depleted stretches of river and reduction in average flow (Anderson et al., 2014). Consequences include reduction in fish species and changes to invertebrate density.

Environmental agencies (usually) insist that a proportion of the river is 'hands off' and cannot be abstracted. At the start of this project, the Scottish Environmental Protection Agency (SEPA) demanded a 'hands off' flow that is equivalent to at least the natural low flow that would, on average, be exceeded for all but 18 days of the year (i.e. Q_{95} or the 5th percentile flow – the flow that would be exceeded 95% of the time). In certain circumstances (e.g. river reaches with populations of salmon and trout, rivers designated as a conservation area, rivers with catchment areas upstream of less than 10km^2 and where the wetted width is significantly reduced at flows below Q_{90}), this must increase to Q_{90} (i.e the flow that would be exceeded 90% of the time or exceeded for all but 36 days of the year). It was not clear what the environmental policy of 'hands off flow' is in China, and hence the SEPA guidelines were adopted, removing Q_{95} from all available flows.

Other SEPA rules were not adopted in this project, including a 'hands off' flow at Q_{80} when the upstream flow reaches Q_{30}, a maximum abstraction of 1.5 times the mean flow and any of the rules regarding protection of flows for upstream fish migration and spawning.

2.8 An introduction to the Yangtze drainage basin

Several large river systems are found within, or at least partly within, China's borders, including four of the ten longest rivers in the world – the Yangtze, Yellow, Ob-Irtysh and Amur-Argun. The Yangtze, at approximately 6,300km long, is the third longest and drains a basin of approximately 1,800,000km^2 (Britannica, 2017) from its source on the Tibetan Plateau to the East China Sea (see figure 2.13), crossing or bordering ten provinces of China. Eight main tributaries join the Yangtze along its length, including the Yalung, Min, Jialing and Han on the left bank, and the Wu, Yuan, Xiang and Gang on the right. In China, the river is known as 'Chang Jiang' or the Long River.

The Yangtze can be sub-divided into three main regions: the upper, middle and lower courses. Originating on the east of the Tibetan Plateau from several tributaries over 5,000m elevation, the Yangtze River runs east across Qinghai province, before turning south through a deep valley at the border of Sichuan province and Tibet before arriving in Yunnan province, and the end of the upper course at Ybin, Sichuan, at approximately 300m elevation. Over the upper course of the river it loses approximately 5200m in elevation over approximately 2600km. Unsurprisingly the upper course is characterised by high mountains and deep valleys.

The middle course of the Yangtze River stretches for approximately 1000km between Yibin in Sichuan province and Yichang in Hubei province, crossing the low-lying Sichuan basin which is surrounded by mountains on all sides. As the river passes into western Hubei province, it flows through the Three Gorges region. The lower course of the river is characterised by extensive plains before eventually reaching the Yangtze delta. DEM representation of the Yangtze is discussed in section 3.3 and an elevation map shown in figure 2.14

Figure 2.13: Location of the Yangtze drainage basin within China (WWF Global, 2017)

Mean annual precipitation in China is greatest in the south-eastern corner i.e. over the middle and lower courses of the Yangtze basin (for more details see section 4.2), with warm, wet

summers and cold, dry winters. Mountainous areas with high precipitation are an ideal hydropower resource. The basin is also home to a significant portion of China's population (approximately 430 million (WWF Global, 2017)), particularly in the eastern two-thirds, with many large cities including Chengdu, Chongqing, Wuhan, Nanjing and Shanghai (see figure 3.25).

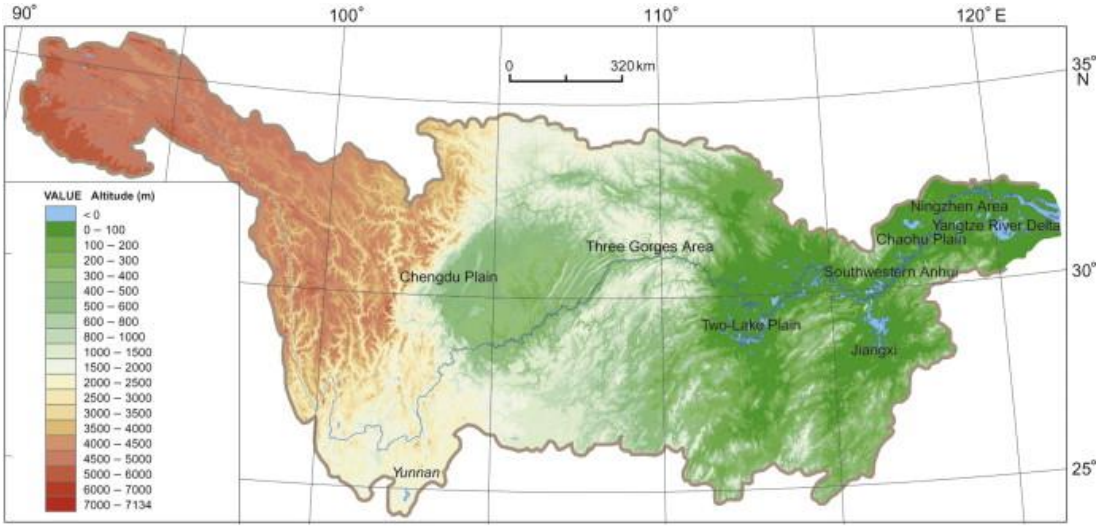


Figure 2.14: Topography of the Yangtze basin [m] (Wu et al., 2012)

2.9 An introduction to hydrology and the hydrological cycle

Hydrology is the study of the movement, distribution and quality of water. Hydropower resource assessment is primarily concerned with surface runoff estimations which are dependent on the hydrological cycle, a central theme of hydrology.

The hydrological cycle (or water cycle), in simple terms, is the transport of water from the Earth's surface to the atmosphere and then back to the Earth's surface, ultimately powered by the sun. Water on the Earth's surface (in oceans, seas and rivers) is heated by the sun, evaporating the water into water vapour. Moisture within soils and on land surfaces is also evaporated and transpired by plants in a combined process known as evapotranspiration. Water vapour rises in the atmosphere due to density differences and buoyancy, and cools due to the lower temperatures of the higher atmosphere, condensing into water droplets. Air currents move the vapour around, and on reaching hilly land, the vapour is forced upwards, which condenses and precipitates (as rain or snow). Water falling on land either runs over the surface or infiltrates the soil where it is stored as soil moisture and eventually becomes

groundwater. Surface water accumulates in streams, rivers and lakes and is ultimately returned to the oceans.

2.9.1 Water balance and hydrological models

Hydrological models of river flow are simplifications of real world hydrological systems, and used to predict the response of a catchment to rainfall, used in both forecasting and hindcasting. A fuller review of hydrological model development, calibration and models that would be appropriate for modelling the Yangtze basin is given in sections 5.1 to 5.3. At their heart, most hydrological models try to maintain a water balance, either at a catchment level or at a cell level for gridded (raster-type) models, which describes the flow into and out of a system.

A general water balance equation re-arranged for surface runoff (R) is given below:

$$R = P - E - \Delta S \qquad (2.2)$$

where R is surface runoff, P is precipitation, E is evapotranspiration and ΔS is the change in storage (i.e. in soil/bedrock or ground water).

Precipitation can also be stored as snow, ice or glaciers and released as meltwater, and therefore a full hydrological model may require input of a snowmelt model, depending upon the location being studied. The quantity of precipitation entering the ground is dependent on the ground type, vegetation, the level of saturation and the rate of precipitation falling. Human intervention may also change the amount of water running over the surface and the flow path (due to man-made surfaces and/or abstraction). A proportion of groundwater may return to the surface through groundwater recharge. Surface runoff moves down the catchment accumulating in valleys to from streams and rivers, eventually reaching the ocean.

2.9.2 Hydrology of the Yangtze

Most of the precipitation within the Yangtze drainage basin is brought by monsoon rains, falling mostly during the summer months (June to August) and often as snow on higher mountains. The middle and lower sections of the basin often experience floods anytime from March/April and over the following eight months. Water levels are usually at a minimum in February. Water volume within the Yangtze mainstream is predominantly contributed by the middle and lower sections (90%).

Mean flow of the Yangtze mainstream by the end of the upper section is already nearly 2,000m^3s^{-1}, increasing to over 5,500m^3s^{-1} after the joining of the first major tributary, the Yalong River. By the time the river has reached the Three Gorges area mean discharge is approximately 24,000m^3s^{-1} and 31,000m^3s^{-1} at the river mouth. This ranks the Yangtze third in the world by volume discharge after the Amazon and the Congo. However, the completion of the Three Gorges Dam has significantly altered the downstream hydrology and ecology (Zhang et al., 2016a). Suspended sediment load at the river mouth is in the order of 480 million tonnes per year.

Flooding occurs due to high precipitation and associated runoff, with flows potentially doubling in peak rainy season, but also due to high sediment load raising the river bed. Since the middle of the 19th century the Yangtze has experienced many catastrophic floods affecting millions of people, particularly in 1931, 1954 and 1998. In 1931 more than 145,000 people (with some estimates as high as 4 million) lost their lives; this was viewed as the deadliest natural disaster of the 20th century. In 1998, 3,704 people lost their lives, 15 million were made homeless and the economic loss was estimated at US$26 billion. Even post the construction of the Three Gorges Dam, heavy rainfall in 2016 caused floods across the Yangtze and Huai rivers, impacting an estimated 32 million people across 26 provinces and killing more than 300.

2.9.3 Hillslope hydrology

Bevan (2012) states ‘hydrological systems are sufficiently complex that each hydrologist will have his or her impression or perceptual model of what is most important in the rainfall-runoff process so that different hydrologists might not necessarily agree about what are the most important processes or the best way of describing them’. Bevan also states that our understanding of hydrological response is still evolving, and a hydrologist’s perception will depend upon experience and the environment studied.

One such perceptual model is given by Bevan: precipitation inputs are not spatially uniform and exhibit rapid changes in intensity and volume over even relatively short distances. This is further interrupted by the vegetation canopy as some of the rainfall will be intercepted and evaporated back into the air. Some of the rain falls directly on the ground as *throughfall* and some will drip from leaves or flow along trunks etc as *stemflow*, which results in local

concentrations of water at high intensities. Snowmelt rates vary with elevation and aspect and the water equivalent of the snowpack can vary dramatically spatially.

On reaching the ground, rainfall or snowmelt begins to infiltrate the soil surface, except over impermeable bare rock, frozen ground and some man-made surfaces, in which case surface runoff will begin immediately. The amount of infiltration will be limited by the intensity of rainfall and the *infiltration capacity* of the soil. If the input rate exceeds the infiltration capacity, *infiltration excess overland flow* will be generated. Soils are heterogeneous in their characteristics and infiltration capacities vary greatly. In some places rainfall rarely exceeds the infiltration capacity unless the ground is saturated from previous events. Bare soil areas are vulnerable to infiltration excess runoff as raindrops can rearrange the soil particles to form a crust. Vegetation on the other hand protects the surface and creates root channels through the soil. Vegetation also controls the degree to which soils freeze. The underlying soil structure and macroporosity of soil control infiltration rates.

Overland flow may also occur due to *saturation excess* mechanisms, usually occurring first in valley bottom areas, areas with thin soils, low permeability soils and flatter areas. Surface runoff is also generated due to return flow of subsurface water. As saturation begins to build at the base of the soil over impermeable bedrock, it will start to flow downwards as *subsurface stormflow*. Some bedrock is penetrable, and water can percolate through the rocks to form *groundwater*.

High on the catchment, surface runoff accumulates to form streams which flow into valleys to form rivers. Slower groundwater can take days or weeks to move down the catchment and return to the surface and hence is a much more consistent contribution to river flow – hence rivers have periods of *baseflow* when precipitation is low and short periods of time of peak flow when heavy precipitation is experienced. River flow is usually measured on a hydrograph, showing the rate of flow, or discharge, versus time at a specific point of a river.

2.10 Meteorological datasets

From equation 2.2, high quality meteorological datasets at sufficiently high-resolution are important inputs into a runoff model. A review of development of precipitation and temperature datasets is given in sections 4.2.1/4.2.2 and 4.3.1 respectively. Available precipitation and temperature datasets and selection for this project are discussed in 4.2.3 and 4.3.2 respectively. Importantly there are limited Chinese sourced datasets and of those few of

those are available in the public domain. Furthermore, reviews of available datasets do not suggest that Chinese datasets would be of higher quality (Yin et al., 2014).

A more detailed introduction to evapotranspiration is given in section 4.1. Evaporation is often measured by the 'pan evaporation' method which measures the water loss from a standardised cylinder. However, the pan method only closely simulates water loss from open water bodies and not transpiration from vegetation. Another manual method is the use of a soil lysimeter, where the weight of a soil column is continuously monitored. Eddy covariance is possibly the most direct method of measuring evapotranspiration (Baldocchi, 2014; Burba, 2013).

More often hydrometeorological equations are used to calculate evapotranspiration, particularly when assessing evapotranspiration over wide areas, of which the Food and Agriculture Organization of the United Nations (FAO) methodology to calculate daily evapotranspiration rates (Allen et al., 1998) is viewed as the most complete. However, the FAO-56 Penman-Monteith method can be difficult to implement due to the lack of availability, or quality, of measured meteorological variables. Energy balance methods are also used increasingly, and a full review of available models and comparison to the FAO method is given in Valipour et al. (2017).

There is increased measurement of evapotranspiration from remote sensing satellites (e.g. Zhang et al., 2016b; Cammalleri et al., 2014; Peng et al., 2016; Majozi, et al., 2017).

2.11 Geographical information systems

Geographical information systems (GIS) are used to analyse, manipulate and present spatial and geographic data, an important aspect of this project. GIS software tools fall into two broad camps: dedicated menu driven GIS systems with a graphical user interface (GUI) or customised GIS code enabled within standard programming languages such as R, Python, C++ and FORTRAN.

ArcGIS developed by ESRI is the GUI type commercial market leader and used extensively in this project in preparation and presentation of data for the R algorithms. ArcHydro is a powerful addition to ArcGIS used to delineate and characterise watersheds in raster and vector formats (see section 2.11.1), define and analyse hydro-geometric networks, manage time-

series data and configure and export data to numerical models (ESRI, 2017). ArcGIS Spatial Analyst adds a range of spatial modelling and analysis tools to the base system.

Other, non-commercial and open source, GIS GUI type systems are available including Q-GIS which provides many of the features of ArcGIS. The Geographical Resources Analysis Support System (GRASS) system is also free and open source but commands are executed via a Unix shell environment. GRASS can also be manipulated through a more traditional GUI such as QGIS.

However, such GIS systems are limited compared to customised GIS code developed through a standardised programming language which can manipulate multiple processors, a necessity when analysing intensive code over wide spatial areas. Unlike GUI type GIS systems, customised code written in R, C++ or Python (for example) can perform computations in parallel, where one 'master' can harness the power of multiple cores and processors at the same time and yet manage the results centrally. There are many libraries available for each of the codes mentioned that enable most of the features of even the most powerful desktop GIS systems. One such essential underpinning library is the Geospatial Data Abstraction Library (GDAL) (initially released by Warmerdam, 2000) which enables multiple codes and GIS systems (including R and ArcGIS) to read and write various GIS formats. Another essential library is GEOS (Geometry Engine Open Source) which delineates and implements standards-based geometry classes available in GDAL.

2.11.1 Rasters and vector graphics

Rasters are gridded rectangular arrays of cells, usually associated with geographical coordinates when used in a GIS context, and each cell is assigned a value (e.g. elevation or precipitation). The size of each cell depends upon the resolution, with each cell representing an area, rather than a point, and the assigned value is the mean of that area.

Vector graphics is the use of polygons to represent images, consisting of a series of nodes or vectors, with interconnecting lines. Each node is given a spatial location. Therefore, when zooming in on vector graphics the image remains sharp and its accuracy depends on the number of nodes initially defined.

2.12 Problem statement and methodology

To search for economical ROR hydropower resource across a large catchment, using the Yangtze drainage basin as a test case, it is first proposed to identify key datasets required to build 'virtual' hydropower stations (see chapter 3). This includes selecting an appropriate digital elevation model and a method to isolate the catchment in question. Other datasets required include locations of lakes and reservoirs, the existing road network, an electrical transmission network and a power production costing model.

Flow is another important variable in determining the hydropower potential of a scheme, which varies daily across the river network. Therefore, it is proposed to develop a high-resolution gridded hydrological model based on the G2G model which was first used to simulate surface runoff flows across England (Bell, 2007a) (see chapter 5). Ultimately, the aim is to generate FDC data for every grid cell across the river network, which requires long time-series of surface runoff data.

Meteorological data is required as inputs to the hydrological model including precipitation, temperature and evapotranspiration data. Precipitation and temperature data was selected from the APHRODITE suite (Yagatai et al., 2012) due to the high-resolution coverage of the Yangtze area at a daily time-scale. Daily evapotranspiration data was developed by using meteorological data available within the MERRA suite (Rienecker, 2012), including net radiation, mean, minimum and maximum daily temperature, windspeed, atmospheric pressure and specific humidity. As the highland regions of the Yangtze drainage basin are cold enough for snow, particularly during the winter, a snowmelt model will be included based on SNOW-17 (Anderson, 1973; Anderson, 2006). For further details of the development of the meteorological datasets see chapter 4.

The hydropower search algorithm development is described in chapter 6. The aim is to scan the catchment river network raster (see section 2.11.1), produced from the hydrological model, looking for cells with sufficient flow. If a flow threshold has been reached, a virtual penstock is extended to the next cell in the flow direction of the river, and if a minimum head threshold is reached, a turbine map is interrogated to find available turbine types. Based on RETScreen's equations (RETScreen, 2004), the virtual power station will be costed and the plant physical and economic characteristics will be saved to a datafile. For each turbine type, the design flow (maximum flow) of the scheme will be iteratively tested. At the end of the

iterative tests, the penstock will be extended to the next cell along the river network and the process repeated until a maximum penstock length is reached.

Finally, those plants that are deemed viable i.e. have a sufficient internal rate of return, will be extracted from the list of trial configurations. A hydropower conflict algorithm will then be run to select the sites with best NPV and remove sites that conflict (i.e. the intake or turbine is between the intake and turbine of another station). This will enable presentation of a final optimised conflict-free network of hydropower stations.

Although different in its implementation, this work builds on Duncan's (2014) study of hydropower resource within Scotland's catchments, which itself builds upon a GIS method to assess micro-hydro projects in Scotland (Forrest, 2006).

2.13 Chapter conclusions

- Global primary energy consumption is rising driven primarily through population growth and an increase in global GDP per capita, dominated by fossil fuel use. This is particularly noticeable in developing countries including China.
- There is overwhelming evidence that the use of fossil fuels for energy is linked to climate change, through anthropogenic emissions of greenhouse gases. This is having a direct impact on human life and natural ecosystems.
- It is predicted that climate change will have a direct impact on hydropower generation.
- Hydropower is a well-established, high efficiency technology. Run-of-river schemes are considered more environmentally benign than large reservoir scheme but rely on the flow at any time which can vary dramatically.
- There are several high efficiency turbine technologies available.
- Hydropower contributes 16% of global electricity production, and China is the leading country in terms of its hydropower capacity.
- There have been many systems-based studies of hydropower potential of a catchment or region, largely calculating the potential of a river/slope rather than costing of virtual power stations. Duncan (2013) trialled virtual power stations across a large region (Scotland) and hence will form an important contribution to this study of a large global catchment.

- Hydropower is capital intensive but has low LCOE compared to other technologies and there is financial risk including reduced future runoff, environmental issues and underestimation of costs.
- There are environmental implications of both traditional dam and ROR hydropower, but these will not be considered as part of this study.
- The Yangtze is the third longest river in the world, and hence the YDB is a particularly large catchment globally.
- In developing tools to understand the flow regime of a rivers it is necessary to understand hillslope hydrology, meteorological datasets and GIS.

3 Methodology: Input data requirements for a hydropower search and mapping algorithm

3.1 Introduction

Searching for and mapping economically viable hydropower sites across the Yangtze drainage basin (YDB) requires development (or appliance) of a hydropower search algorithm. Before developing an appropriate algorithm, it is necessary to understand the input data requirements and ensure these are available. This chapter considers the required inputs to a hydropower scheme assessment model and selection of appropriate GIS input data available in the public domain. GIS spatial data is increasingly available in gridded raster or vector format covering much of the globe and readily integrated within computer-based models.

3.2 Basic information required in assessing a hydropower scheme

Although hydropower schemes vary significantly in their size, choice of equipment and implementation, they all have common attributes. Figure 2.9 is a pictorial representation of a typical ROR installation. A proportion of a river is diverted into a penstock (pipe) at the weir intake, which due to sufficient head has enough pressure to turn a turbine and generate electrical energy in the power house. This electrical energy is transmitted to an electrical grid network (possibly off-grid) and the harvested water discharged back to the river. Depending on local environmental rules and the design of the scheme, a proportion of the river flow must remain within the river. Large dam projects, although more complex, have similar attributes with the addition of water stored behind the dam in a reservoir.

The equation for the power generated from water flowing through a turbine is:

$$P = \rho g H_{net} Q \eta_t \eta_e \quad (3.1)$$

where P represents power [W], , ρ is density of water [kgm^{-3}], g is the acceleration due to gravity [ms^{-2}], H_{net} is the net head [m] after considering pipe losses due to friction and discharge losses, Q is the flow through the turbine [m^3s^{-1}], η_t is the turbine efficiency and η_e is the electrical efficiency. A developer normally considers the electrical efficiency, η_e, to be losses within the generator and sometimes the transformer, but in this exercise losses are also considered within the electrical distribution network (which is not normally owned by the developer).

From the above it is clear that some basic information is necessary when searching for viable hydropower schemes to input into a power production and costing model. When testing whether a point on a river is suitable for hydropower it is necessary to have access to the following datasets:

- H_{net} is the difference in height between the intake and discharge points of the flow. Hence **elevation data** relating to the intake and discharge is required.
- Intake and discharge points must be on rivers which requires a **spatial layout of the river network**.
- Losses in head are due to friction losses in the penstock, which depends on the flow through the penstock, length of the penstock and materials chosen. The laying of the penstock needs to consider the terrain in the region of the river and hence terrain data (**elevation dataset**) and a **water body dataset** (i.e. lakes/reservoirs) are required.
- Rivers vary in their flow over a year due to variations in meteorology and hence an understanding of **river flow**, Q, is important, which is also used to select an appropriate turbine and in modelling of turbine efficiencies.
- A weir dams a river to enable a proportion of the flow to be diverted into the penstock. **River width and depth** is used when attempting to cost weir construction.
- Electrical efficiency, η_e, is in part due to transmission distances and transmission lines and needs to be costed. An understanding of the **electrical grid network** is also required.
- Construction of hydropower requires access by road. Costing of necessary new roads requires an understanding of the **original road network**.

All these datasets need to feed into a robust power production and costing model. Other data is necessary for these models but in this chapter the focus is on selecting or developing GIS datasets. An overview on how these datasets feed into a power production and costing model is given in figure 3.1.

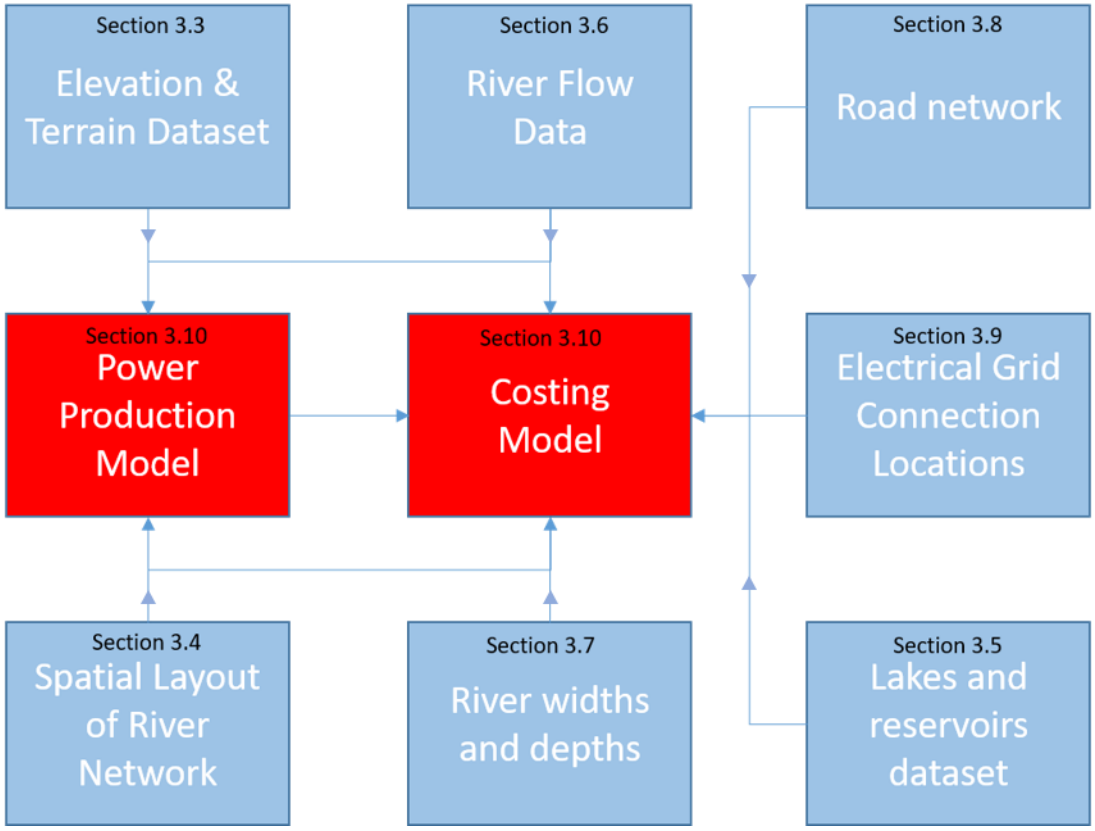


Figure 3.1: Inputs required for developing a hydropower search algorithm

3.3 Selection of an elevation dataset

Elevation data of the Earth's surface, or at least parts of it, has long been available in the form of traditional maps conducted by land survey, which vary in quality. Although potentially useful if investigating an individual hydropower site, they have limited use if trying to automate a hydropower search over a wide area. Of more use is elevation data in a digitised form known as a digital elevation model (DEM) – a 3D representation of a terrain's surface. DEMs are available in the public domain or for purchase, and usually presented in the form of a gridded raster - a grid of cells or a heightmap. DEM is an all-encompassing term which includes digital surface models (DSM) and digital terrain models (DTM), where DSMs

represent the Earth's surface with objects on it (i.e. buildings/vegetation) whereas a DTM represents the elevations of the bare ground, the latter most useful in hydrology and hydropower assessment. DEMs in the form of a raster are relatively easy to manipulate in code and therefore useful for this study.

Access to highly accurate DEMs, particularly of mountainous regions, has been restricted until recent times due to difficulties including: partial area of coverage, high costs of development and physical restrictions in conducting traditional land survey (Mispan et al., 2015). However, by utilising satellite technology much of the Earth's surface has now been digitised and is available to download. Online mapping tools such as Google Maps and Google Earth (Google, 2017) are highly detailed and at high-resolution (the latter is approximately 15m per pixel) but the proprietary code makes them difficult to manipulate in computer based models.

High-resolution DEM datasets representing China have been produced by government mapping agencies but are not publicly available, and in this study a DEM available in the public domain was sought as the techniques developed could then be used to study other global catchments. Global DEMs are preferable as, before analysis, it is unknown whether drainage basin boundaries of a river are wholly within one country. Commercial datasets at very high-resolution are available for purchase including WorldDEM^{TM} at 12m x 12m (Airbus Defence and Space, 2016) and AW3D World 3D Topographic Data, a 5m gridded DEM (NTT Data Corporation, 2016). Global 30-arcsecond (approximately 1km grid square) Elevation data (GTOPO30) (Gesch et al. 1999) is free to use for non-commercial organisations but is of variable quality and has been superseded by two higher resolution global satellite-generated DEM products - the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (METI (Japan) and NASA (U.S.A.), 2011) and the Shuttle Radar Topography Mission (SRTM) (Farr et al., 2007). Both are now available at 1-arcsecond DEM (approximately 30m), covering most of the globe.

3.3.1 ASTER vs SRTM DEM for spot height evaluation

The ASTER Global DEM was generated using stereo-pair images collected by the ASTER instrument on board the TERRA satellite, with coverage from 83 degrees north latitude to 83 degrees south, encompassing 99 percent of the Earth's landmass (see figure 3.2). Improvements to the dataset were made in V2 (2011) with the addition of 260,000 additional stereo-pairs increasing coverage, accuracy, spatial resolution and reduction of artefacts,

however still with the warning that anomalies and artefacts exist. Artefacts are errors in the perception and representation of information introduced by the equipment and techniques employed, particularly relevant to signal processing.

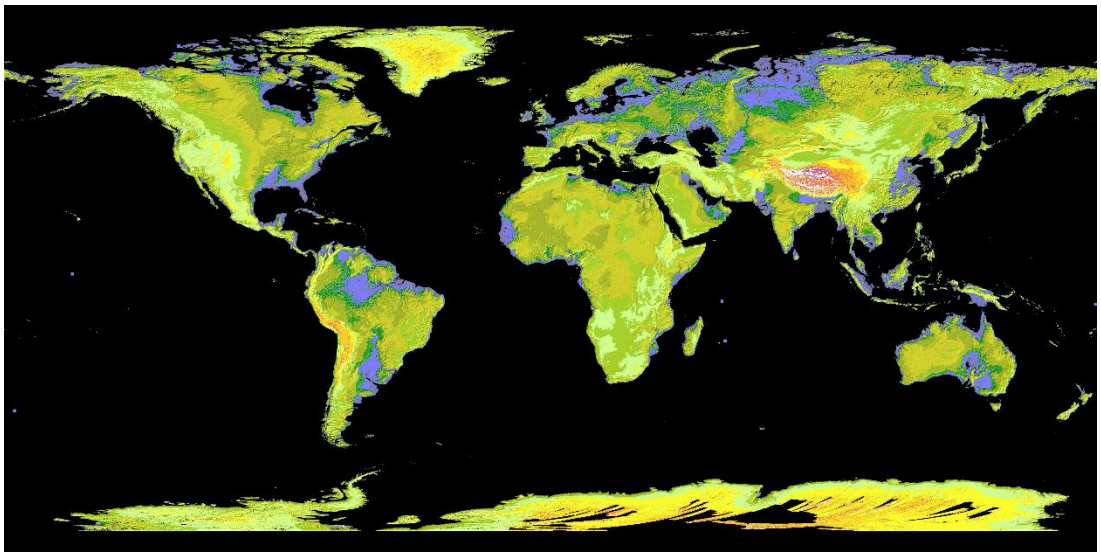


Figure 3.2: Coverage of Version 2 of the ASTER Global DEM (GDEM) (NASA Jet Propulsion Laboratory, 2011)

SRTM data employed synthetic aperture radar (SAR) interferometry using two radar images on board the Space Shuttle Endeavour with coverage between 60 degrees north latitude to 56 degrees south latitude (see figure 3.3). Validation of the dataset revealed accuracy better than 9m with greatest error over steep terrain (Himalayas, Andes etc) and very smooth sandy surfaces. Voids in the data are present due to steep slopes facing away from the radar and smooth surfaces such as water and sand which reflect the radar image. SRTM measures an effective height as the radar could not penetrate dense vegetation canopies and man-made structures, and yet SRTM radar could penetrate frozen snow/ice and very dry soil further leading to error. At the start of this project, SRTM data outside of the United States was only available at 3-arcseconds (approximately 90m) but this has subsequently been upgraded to 1-arcsecond (2015).

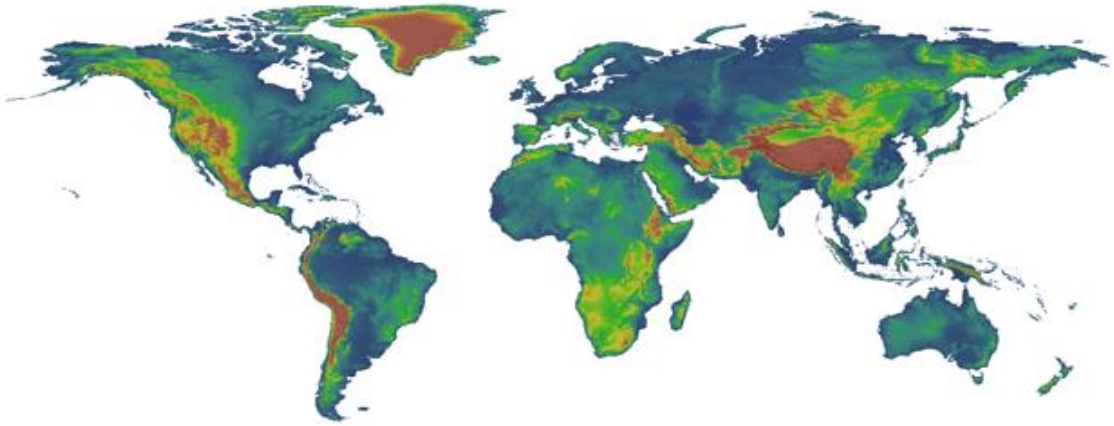


Figure 3.3: Coverage of v4.1 of the SRTM database (CGIAR, 2008)

Many studies have compared ASTER data to SRTM data. A study of volcanic features in Hawaii and Tanzania demonstrated that SRTM-3s data had a higher vertical accuracy when compared to ASTER-1s (10m vs 13m root mean square error) but a greater underestimate of cone heights due to the averaging effect of decreasing spatial resolution (Kervyn et al., 2008). This increased accuracy in SRTM data is also reflected in studies of the Australian continent (Rexer & Hirt, 2014), assessment of highland areas of peninsular Malaysia (Mispan et al., 2015) and in a study of North East Tunisia (Ouerghi et al., 2015). More relevant to this project, studies comparing SRTM vs ASTER over areas of China also reveal higher accuracy of the SRTM dataset including a comparison over Zhejiang (SE China) (Jing et al., 2014) and in an evaluation in South East Tibet (Wang et al., 2011). However, most of these studies conclude that either model is suitable, particularly if one or other of the models contains a high number of void data in the area being studied.

If used for purely spot height purposes, the lower resolution but higher accuracy SRTM-3s data is a more appropriate choice, as lower resolution models are easier to manipulate in code due to the lower data processing requirements. However, a DEM is also required to derive the river spatial network and therefore voids and anomalies within the data are also important to consider.

3.3.2 DEM selection suitable for watershed delineation and derivation of river networks

Even with limited knowledge of the meteorology of a catchment, a spatial layout of the river network and drainage basin boundaries can be extracted from an accurate DEM on the basic assumption that water flows downhill in the steepest downhill direction, accumulating in valleys and channels. This methodology will be discussed in detail in section 3.4. Comparison of ASTER and SRTM DEMs for watershed delineation in Ecuador showed the area of a catchment derived using SRTM data was closer than that derived using ASTER data to the area of the catchment delineated by hand using accurate topographical maps (Pyrde et al., 2007). In determining the reasons for the greater discrepancy, it was found that the ASTER data had a greater number of 'fills' of void and anomaly data. Hence, large inaccuracies and missing data (voids) can disrupt river networks and catchment boundary delineation and hence it is important to find a DEM suitable for hydrological modelling i.e. void filled. Both ASTER and SRTM original data contain such anomalies making them both inappropriate for hydrological modelling in their raw form.

3.3.3 Selection of a void-filled DEM

From the above, a void filled SRTM dataset would be desirable as an underlying DEM for both river network delineation and spot height measurement. There are a number of void-filled DEMs available, but some research groups have only used interpolation from surrounding data within the set, which is potentially prone to large error. NASA released a void-filled SRTM DEM known as SRTM Plus or SRTM NASA Version 3 (NASA 2015), with voids mostly filled using good ASTER data and the US Geological Survey's GMTED2010 model under NASA's "Making Earth System Data Records for Use in Research Environments" (MEaSUREs) program. The Consultative Group for International Agricultural Research Consortium for Spatial Information (CGIAR-CSI) have produced a 3-arcsecond void-filled SRTM based DEM for the entire world (Jarvis et al., 2008), which when produced the CGIAR were confident that it was the highest quality SRTM dataset available.

In 2008 the World-Wide Fund for Nature (WWF) in partnership with the USGS and other organisations released the HydroSHEDS (**Hydro**logical data and maps based on **SH**uttle **E**levation **D**erivatives at multiple **S**cales) suite of products (see figure 3.4), including DEMs specifically aimed at providing global-scale hydrographic information (Lehner et al., 2008).

Included in the suite are geo-referenced datasets in both vector and raster format. These include conditioned DEMs, flow direction rasters and flow accumulation rasters at 3-arcsecond, 15-arcsecond (approximately 450m at the equator) and 30-arcsecond (approximately 1km at the equator) resolution and river network and drainage basin SHAPE files at 15-arcsecond and 30-arcsecond resolution. The shapefile format is a geospatial vector data format which can be interpreted by GIS software, developed and regulated by ESRI.

Preliminary assessment of HydroSHEDS (by the HydroSHEDS development team) indicates the quality significantly exceeds existing global watershed and river maps. The accuracy of HydroSHEDS is declared better than the HYDRO1k data, Digital Chart of the World (DCW) data (ESRI, 1993) and the river layer in ArcWorld (ESRI, 2015). However, HydroSHEDS derived river networks are considered less accurate than high-resolution local river networks as depicted in local maps and remote sensing imagery. Other errors in the data will be considered in the next section. Despite this, as the HydroSHEDS suite has been designed specifically for hydrological applications, has a range of resolutions to work with and is considered void-free and high quality, the HydroSHEDS DEM was selected as the underlying DEM for this project.

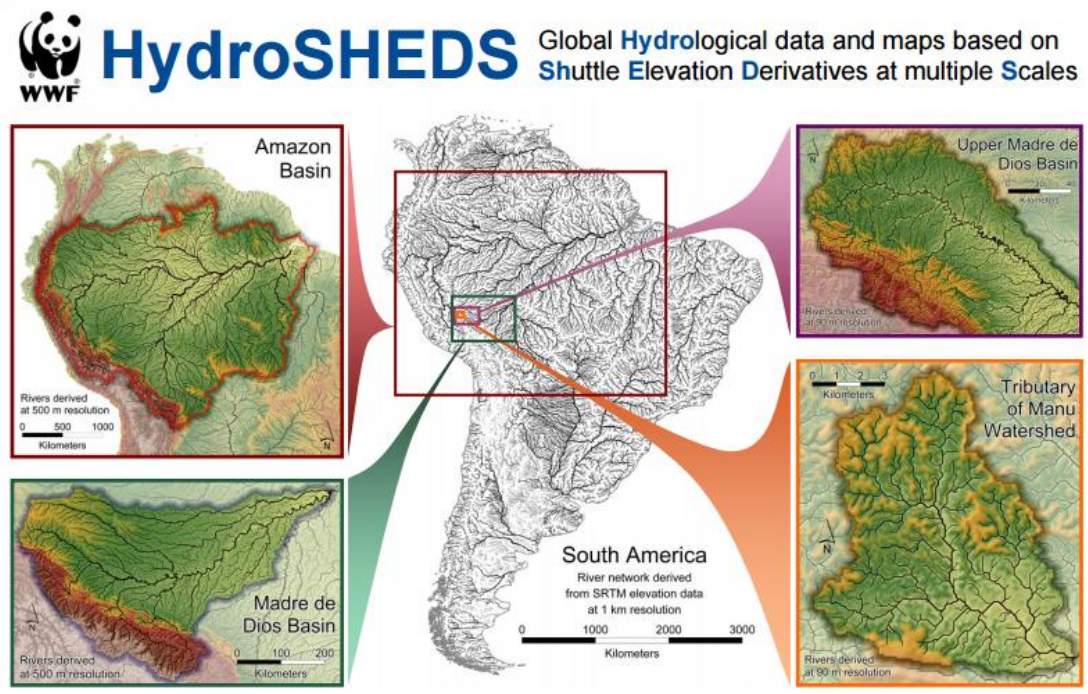


Figure 3.4: Leaflet of HydroSHEDS showing drainage basin and river networks derived within the Amazon Basin, South America (Lehner et al., 2008)

3.3.4 An Overview of the HydroSHEDS data suite

The goal of HydroSHEDS was to generate data layers to support regional and global watershed analyses, hydrological modelling and freshwater conservation planning at resolutions that have previously been unachievable, and free for non-commercial use. The flow diagram (figure 3.5) shows the development of the HydroSHEDS suite of products (by the HydroSHEDS authors) and will be explained in subsequent sections.

Geospatial datasets vary in their accuracy, resolution, method of production and errors, and these characteristics may render a dataset unsuitable for use in certain applications, but make it perfectly acceptable in others. It is therefore important when selecting a DEM for both spot height measurement and hydrological modelling to understand how the dataset was developed. The following is a summary of the development methodology of the HydroSHEDS products. It is important to note that the following notes on the HydroSHEDS suite development (sections 3.3.4.1 to 3.3.4.7) outline the steps made by the authors/developers of HydroSHEDS and not the author of this work, and the level of detail is included for the following two reasons; 1) This explains why HydroSHEDS is suitable for hydrological modelling; 2) Normally when selecting a DEM for hydrological modelling there are number of preparatory steps required, which in this case were not necessary as they were already carried out by the HydroSHEDS development team.

3.3.4.1 Data sources of the HydroSHEDS data suite

Data sources contributing to the HydroSHEDS data suite primarily included SRTM V1 unfinished data at both 1-arcsecond (SRTM-1) and 3-arcsecond (SRTM-3) resolution. SRTM-3 data was generated by averaging of the SRTM-1 data i.e. the height of one pixel of SRTM-3 data is the average of the 3x3 SRTM-1 kernel.

Further sources include SRTM V2 finished data with quality assurance steps conducted by the National Geospatial-Intelligence Agency of the U.S. Department of Defense (NGA) to meet data standards of the Digital Terrain Elevation Data (DTED®) format (NGA, 2017). Quality assurance steps included filling small voids by interpolation but leaving large voids in the data. Ocean elevations were set to 0 metres and lakes of greater than 600m in length were flattened and set to a constant height. Rivers of more than 183 meters in width were delineated and monotonically stepped down in height i.e. to ensure rivers always flowed downhill. Islands were depicted if they had a major axis exceeding 300 meters or the relief was greater than 15

meters. Finishing steps were performed at the original 1-arcsecond resolution, resulting in DTED Level 2 data (DTED-2) products, and DTED-2 was then aggregated into 3-arcsecond DTED-1 data by sub-sampling where the height of the centre pixel of the higher resolution data is assigned to the 3x3 kernel. The SRTM and DTED elevation data was processed and mosaicked into 15,000 one-degree by one-degree tiles, available to download via anonymous ftp.

Water body data sources were also used in the dataset generation. A by-product of the data editing to produce the SRTM DTED-2 data are SRTM Water Body Data (SWBD) vector shapefiles depicting ocean, lake and river shorelines. Landcover water layer, mostly derived from Landsat 5 data, and medium-scale maps were also used as supplemental data (NASA/NGA, 2003). The Landsat 5 data was collected a decade earlier and hence the depiction of water in the SRTM data showed significant seasonal and temporal differences compared to ancillary sources. Shoreline, lake and reservoir polygons within the Global Lakes and Wetlands Database (GLWD) (Lehner and Döll, 2004) were also used as source data, at a scale between 1: 1,000,000 to 1: 3,000,000.

Source river network data included DCW, a global vector map at 1: 1,000,000 resolution which at the time was generally considered the most comprehensive and consistent global river network data available. ArcWorld global vectorised river network was also used at a scale of 1:3,000,000 which distinguishes rivers into natural and artificial and, in some areas, is more accurate than DCW despite its coarser resolution, particularly regarding major rivers and lakes.

3.3.4.2 HydroSHEDS dataset development – Combination of SRTM and DTED data

The SRTM-3 and DTED-1 versions of the SRTM data showed both advantages and disadvantages for hydrological applications. SRTM-3 data used an ‘averaging’ methodology which reduces high frequency noise in the data, a characteristic of radar-derived elevation data and is preferred over sub-sampling by the research community (NASA, 2015). However, SRTM-3 data poorly represents open water surfaces and shorelines and DTED-1 has been specifically corrected to represent these features, although the correction procedure introduced artefacts relevant to hydrological applications. Therefore, both SRTM-3 and DTED-1 data were used in combination using the minimum pixel value of either to generate the initial HydroSHEDS model, which is desirable for later identification of drainage basins.

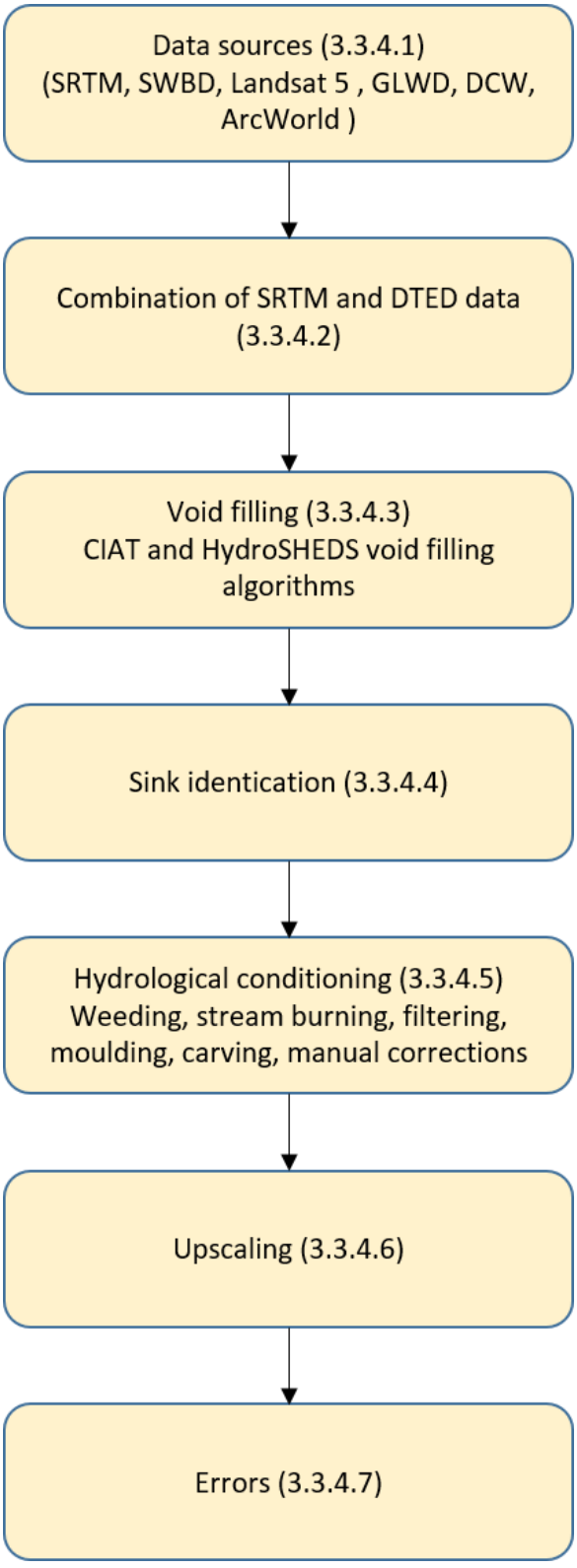


Figure 3.5: Flow-chart of HydoSHEDS data suite development

Both ocean surfaces and land near the coast can have elevation values of 0 or negative and hence using elevation alone does not allow for clean identification of ocean shoreline. SWBD was therefore employed as ancillary data and ocean areas in the HydroSHEDS model reclassified as no-data. To remove small artefacts, the land surface was then extended by one pixel and the boundary smoothed. Some large rivers in the SWBD dataset extend far off land and therefore to correct this, the shoreline was modified by the DCW dataset. When upscaling HydroSHEDS to lower resolution datasets, any cell containing at least one 3-arcsecond land cell was classed as land.

In both SRTM-3 and DTED-1 data, the original 1-degree by 1-degree tiles are identified by the co-ordinates of the centre of the lower-left pixel, resulting in overlap of adjacent tiles and introduction of artefacts when aggregating to lower resolutions. For this reason, the original data was shifted by 1.5-arcseconds to the north and east. In deriving river networks this data shift can be considered negligible, but a user must be aware if comparing the HydroSHEDS data to the original SRTM data.

3.3.4.3 HydroSHEDS dataset development – Void filling

As discussed earlier, the SRTM data contains no-data (void) areas due to radar specific problems preventing the production of reliable data, particularly over water bodies, in mountainous regions such as the Himalayas and Andes due to the radar shadow effect and bright bare sand and rocky areas such as the Sahara Desert. Voids can cause significant problems in deriving hydrological products. Two different void filling algorithms have been employed for HydroSHEDS: CIAT (Jarvis et al. 2004) and a method specifically developed for HydroSHEDS, both used in combination.

The CIAT methodology fills voids using an interpolative technique where SRTM data is used to produce contours at 10m intervals and interpolated using TOPOGRID (ArcGIS) based on algorithms by Hutchinson (Hutchinson, 1988; 1989). This produces a smooth elevation surface and captures macro-scale features well if the voids are small to intermediate in size. The HydroSHEDS void-filling algorithm employs an iterative neighbourhood analysis, particularly applicable to large rivers and lakes. Both algorithms were employed and the minimum pixel value after employing CIAT and the HydroSHEDS algorithm used, unless the HydroSHEDS pixel was 30m lower than CIAT, in which case the pixel value is set to CIAT minus 30m.

In large void areas, entire mountains can be lost to the void-filling algorithm and hence starting from 0.03 degrees from the rim of the void, elevation values were inserted from GTOPO30. The filled voids were merged into the initial HydroSHEDS data providing a continuous elevation surface with no void regions.

3.3.4.4 HydroSHEDS dataset development – Sink identification

Sinks are single or multiple cells (pixels) surrounded by higher elevation land which can occur naturally representing endorheic (inland) basins with no outlet to the ocean. However, in most cases, sinks in the data are erroneous and disrupt continuous flow across a DEM in hydrological applications. Hence sinks are typically removed before deriving a river network by raising the elevation values within the sink until an outflow is generated.

Any sink deeper than 10m and larger than 10km^2 was considered potential to be a natural sink and, after comparing to other digital datasets and maps/atlases, several thousand natural sinks were identified across the globe. However, identifying natural sinks is subjective, difficult and ambiguous.

3.3.4.5 HydroSHEDS dataset development – Hydrological conditioning

Original DEMs have other characteristics making them unsuitable for hydrological applications. One such characteristic is the influence of radar heights influenced by vegetation cover which, particularly in areas of low relief, cause significant errors in derived river courses and flow directions. A series of GIS processes are routinely applied to DEMs used in hydrology, and the hydrological conditioning of the HydroSHEDS dataset was aimed to compromise between forcing the DEM to produce correct river network topology and preserving the original SRTM data. The editing of the original data may make it unsuitable for other applications.

Identified rivers and lakes that matched those within SWBD were deepened by 10m to ensure water remains within these features. Void filled areas were also deepened as these may represent open water bodies. In coastal areas, mangrove and vegetation belts may appear as a low embankment blocking direct flow to the ocean. Therefore, a region 0.02 degree wide at the coast was 'weeded' by reducing every third random cell by 5 metres to force breakthroughs in the false embankment. All cells in the location of perennial and intermittent rivers and lakes of ArcWorld and GLWD were 'stream burned' – i.e. the elevation lowered – with a buffer of

0.005 degrees around the river course. Streams were burned (i.e. deepened) by 12m at the thalweg (lowest point of the river) and in a step-wise manner up to 2m at the buffer edge, whilst lakes were burned to a depth of 14m and a buffer distance of 0.0025 degrees. Stream burning is often used to enforce a river course into a DEM.

Other techniques included filtering by applying a directional 3x3 neighbourhood analysis to remove spikes and wells, to ensure single pixels do not block continuous flow. Valley courses were moulded and deepened by 3-metres, a technique specifically developed for HydroSHEDS to improve river delineations in tropical lowland areas by removing small obstacles. A river map was then derived from the conditioned elevation surface with rivers defined as an upstream catchment of 1000 cells (approximately 8km^2 at the equator) and projected onto the original HydroSHEDS elevation model and any elevation rises along rivers were identified. These rising reaches may represent dams, bridges, embankments or narrow gorges and hence potentially incorrectly identified in the sink-filling algorithm. To minimise this effect, the rising reaches were levelled in the initial elevation data and the hydrological conditioning steps were repeated, hence the rivers were effectively 'carved' through the barriers.

Further manual corrections were employed to correct errors, particularly in flat areas with varying vegetation cover. In some areas, elevation values misrepresent actual flow conditions and hence the burning depth was individually adjusted.

3.3.4.6 HydroSHEDS dataset development – Upscaling

All the above steps were made at 3-arcsecond resolution, but for global or continental use, or in this case modelling of large catchments such as the Yangtze, coarser resolutions may be more desirable. Although fast, upscaling of the elevation data (e.g. by averaging the data) and deriving a new drainage map may produce a river network of low quality due to the significant loss in the original data. A slower, but more accurate methodology, is to produce a river network at high-resolution and upscale this to a lower resolution – preserving the original network. As the latter is complex and difficult to realise on a global scale, a compromise solution was sought. The HydroSHEDS methodology employed was as follows:

1. The Void-filled DEM was upscaled from 3-arcsecond resolution to the desired resolution. An algorithm calculated both the mean and minimum value of each kernel in the aggregation process and the average of these values found, emphasising valleys.

2. A network of rivers was found at 3-arcsecond resolution defined with a minimum catchment of 8km^2 (1000 cells).
3. The rivers were burned into the upscaled elevation surface at a depth defined as the sum of a constant (500m) and a value dependent on the size of the respective river reach (0-400m, proportional to the logarithm of upstream cells). This large burning depth ensured river channels were preserved.
4. Sinks were filled in the upscaled and burned elevation surface and new drainage directions calculated.

It is important to note that this methodology is only used for calculating drainage directions as the elevation data does not represent natural conditions – hence this particular DEM is not available for download. Stream networks derived from the drainage direction map (or flow direction raster (FDR)) should be in close alignment with the original river network.

3.3.4.7 Errors in the HydroSHEDS data

River networks derived from DEM datasets are susceptible to error, particularly in flat regions without well-defined relief, although such areas are generally not considered to be highly suitable to hydropower. Radar derived products are also susceptible to error as they are influenced by vegetation and surface effects including roughness, wetness, low backscatter from open water surfaces and radar shadow (Freeman, 1996). Known errors within the HydroSHEDS datasets include:

- Areas of low or ill-defined relief, including lake surfaces
- Areas of varying vegetation cover due to the radar signal reflecting from atop the vegetation
- Low-relief coastal areas in part due to the barrier effect of mangroves
- Large scale roads and clearings in vegetation
- Rivers less than 90m wide enclosed by riparian vegetation (referring to vegetation on river banks) causing the river channel to be elevated
- Braided rivers and deltas due to the use of a single direction flow algorithm (see 3.5)
- Narrow gorges less than 90m wide as they can appear closed
- Inland sinks and depressions
- Areas of no-data voids in the original SRTM data

3.3.5 Selecting the resolution of the DEM

The development methodology of the HydroSHEDS products make them suitable both for spot height measurement and for modelling of hydrology within a catchment. Ideally the products selected from HydroSHEDS would be at the highest resolution (i.e. 3- arcseconds) but this comes at a computing cost. More data means more accuracy (in this case) but longer computing time. The YDB is reported as approximately 1.8 million km^2 and representing the whole drainage basin at the different available resolutions available would require:

Resolution	Number of cells to represent the Yangtze drainage basin (approx.)
3-arcsecond	222 million
15-arcsecond	9 million
30-arcsecond	2 million

Table 3.1: Approximate number of cells to represent the Yangtze drainage basin (YDB) at different HydroSHEDS resolutions

Table 3.1 shows that a significant number of cells are required to model the YDB at 3-arcsecond resolution and selecting 15-arcseconds reduces that number by approximately 213 million, or a by factor of 25. As the upscaling techniques used in HydroSHEDS preserves the river network (mostly), it was decided to model the YDB using the HydroSHEDS data at 15-arcsecond resolution. This is a compromise between model accuracy and model efficiency.

3.3.6 Final HydroSHEDS DEM dataset selected at 15-arcsecond resolution

HydroSHEDS data was produced on a continental basis and Asia was completed in March 2007. This includes a conditioned DEM raster with elevation in metres and a geographic projection referenced to WGS84 (World Geodetic System 1984) horizontal datum, selected as the DEM for this project and referenced in subsequent notes as HDEM15s. The geographic term indicates it is a latitude/longitude projection meaning the cells change size as they move away from the equator. At the equator (i.e. a latitude of 0^o) 1 degree east-west measures approximately 111km, whereas at 56^o (the approximate latitude of Edinburgh) 1 degree east-west measures approximately 62km. Note that 1 degree measured north-south does not change however far from the equator. Hence in this DEM dataset, cells further away from the equator are smaller in size than cells nearer the equator, which must be factored in to any subsequent modelling.

The Asia HDEM15s is downloadable as a 207MB zip file and shown at a scale of 1:30,000,000 in figure 3.6 (displayed within ArcGIS with a topographical colour scheme representing differences in elevation; brown high and green low). The file is 961MB uncompressed with a grid of 30,000 rows/16,800 columns and 504 million cells.

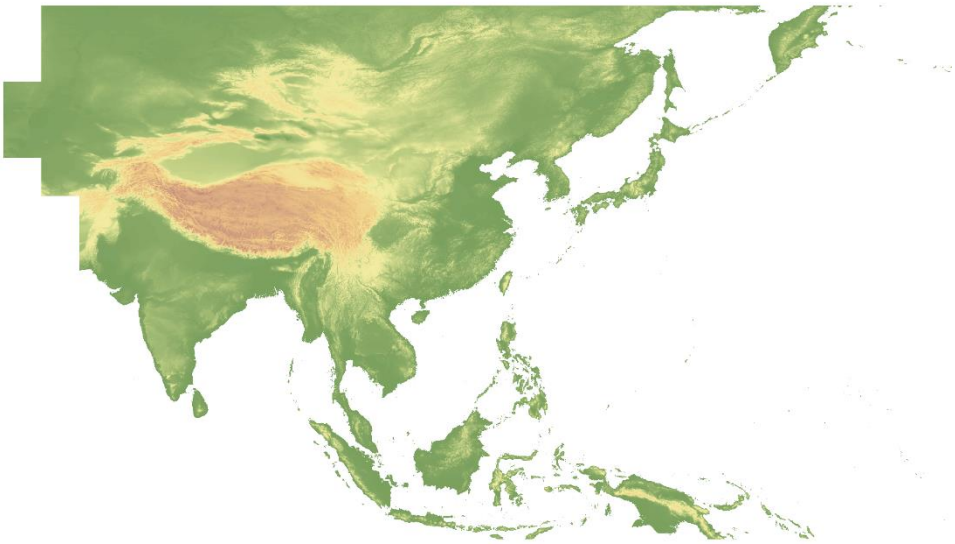


Figure 3.6: HydroSHEDS Asia DEM at 15-arcsecond resolution (HDEM15s) in ArcGIS at a scale of 1: 30,000,000

As the dataset is magnified the individual cells begin to become clear. Figure 3.7 is a mountainous section of the above at a scale of 1: 25,000 and the elevation, longitude and latitude of each cell can be obtained for any individual cell. Validation of the dataset will be discussed in 3.4.9.

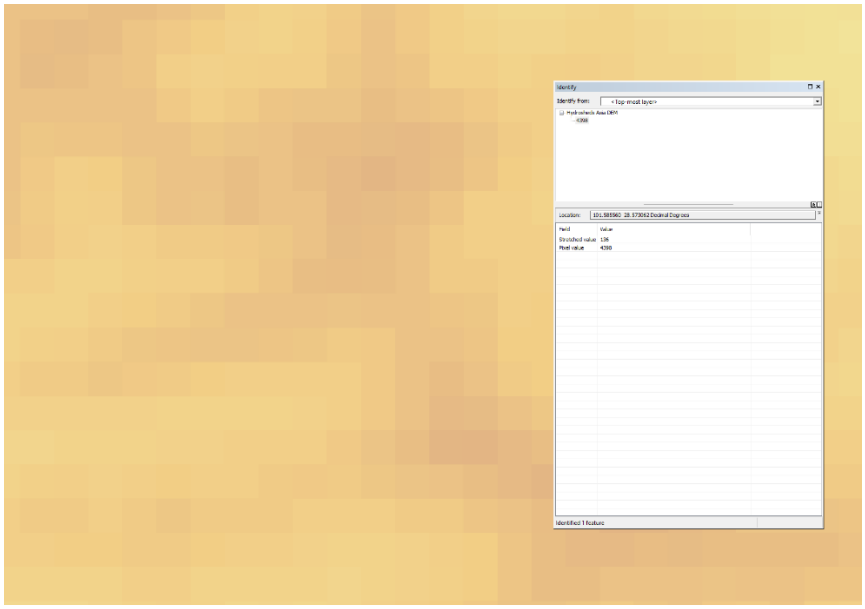


Figure 3.7: HydroSHEDS Asia DEM at 15-arcsecond resolution (HDEM15s) in ArcGIS at a scale of 1: 25,000 showing individual raster cells and feature information of one cell; land area of the above image is approximately 11.56km x 7.86km (90.91km^2)

3.4 Deriving the Yangtze drainage basin and river network

As mentioned in section 3.3, river network and drainage basin boundaries can be derived from a high-quality DEM. This has already been achieved within the HydroSHEDS suite and vector shapefiles are available for download, representing river networks and drainage basins of major rivers (including the Yangtze). However, the HydroSHEDS river network includes rivers with a minimum catchment of 8km^2 and it is possible that rivers of smaller catchment area may be suitable for hydropower. A river network dataset is not only necessary for testing river locations for suitable hydropower sites, but also in modelling flow across a DEM, as the speed of water flow within a river is different than water flow across non-river environments (i.e. over grass/soils). The methodology to derive drainage basins and river networks of smaller upslope catchment size (which was carried out by the author of this work) will be discussed in the following sections and a flow chart of the process shown in figure 3.9.

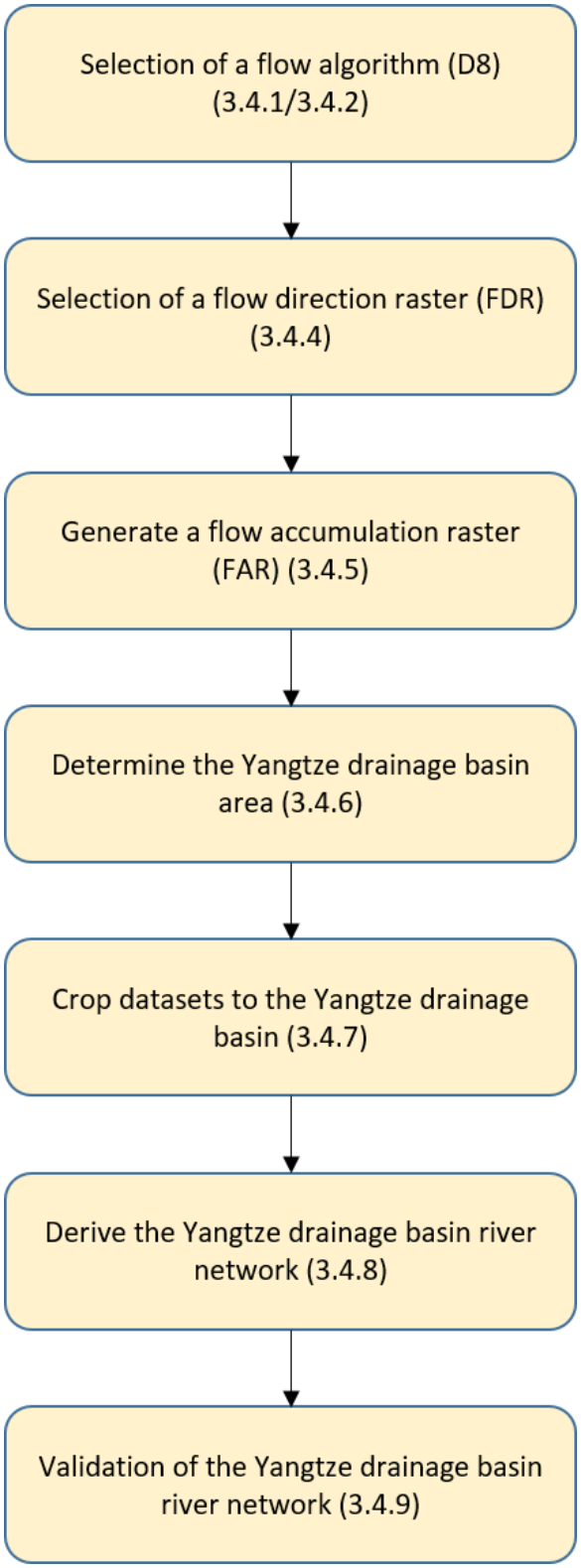


Figure 3.8: Flow diagram of generation of Yangtze drainage basin and river network

3.4.1 The D8 Algorithm and flow direction raster

Automation of drainage network extraction from DEMs is most commonly based on the 'Deterministic 8' or 'D8' algorithm (O'Callaghan and Mark, 1984), using surface water accumulation to define the network. This model is included in tools such as ArcHydro, a plugin to ArcGIS, and the algorithm works by allowing water to flow from a cell to one other cell in only one of the eight cardinal directions.

Raster grids based on squares/rectangles (with four sides, four corners) are generally used in the literature in flow modelling. However, in other fields, such as Lattice Boltzmann modelling, there has been use of geodesic grids based on the subdivision of a polyhedron to subdivide the surface. Geodesic grids have the advantage of being largely isotropic across the surface, can easily increase in resolution by division, have no single points of contact with neighbouring cells and do not suffer from over-sampling near to the poles. However, they are more complicated to implement than rectangular grid rasters, and in this work only rectangular grid rasters are used.

ArcHydro's methodology is based on Jenson and Domingue's (1988) method and includes conditioning of the DEM prior to flow routing. However, this conditioning step has already been completed as part of the HydroSHEDS development. An FDR can be realised by assigning a flow direction to each cell, encoded to correspond to one of eight orientations as shown in figure 3.9.

Figure 3.9: Flow direction coding within ArcHydro (ArcGIS)

Flow directions are encoded in powers of two so that surround conditions correspond to unique values when the encoding is summed for any unique set of neighbours. Figure 3.10 is an example DEM (a) and the corresponding FDR (b) encoded in the ArcGIS format. The directions of flow are shown in (c).

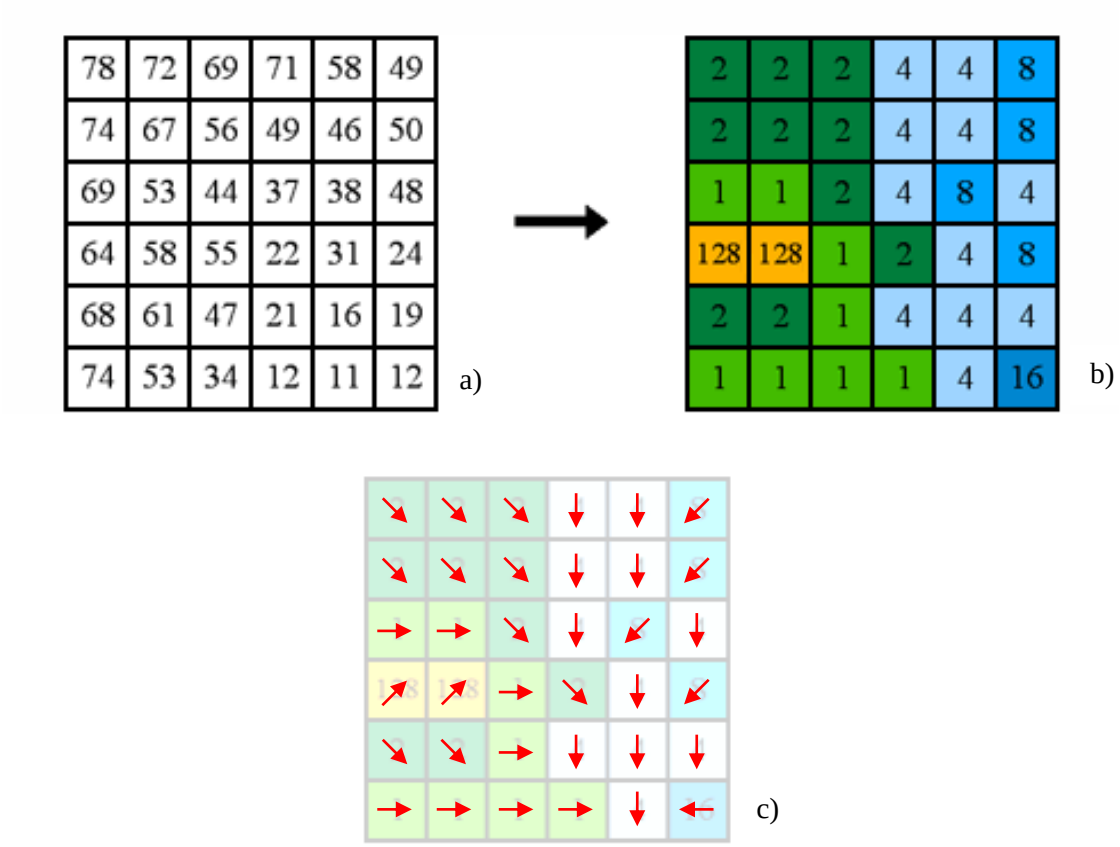


Figure 3.10: An example a) DEM, b) corresponding flow direction raster and c) actual flow directions

Direction of flow is determined by the direction of steepest descent, or maximum drop, from each cell, calculated as follows:

$$Maximum\ drop = \frac{change\ in\ elevation}{distance} \times 100 \qquad (3.2)$$

If two neighbouring cells surrounding the cell of interest have the same elevation, the algorithm will flow to the nearest cell i.e. assuming cells in the diagonal (NE, SE, SW and NW) have a greater distance to travel. If two cells are of equal weighted drops, one is arbitrarily selected. Using this methodology, water flows over the DEM moving towards the catchment outlet or towards a natural (or erroneous) sink. If all neighbouring cells are higher than the centre cell, then the flow is terminated and assigned a negative value.

3.4.2 Alternatives to the 'D8' algorithm

The D8 algorithm is labelled a single-flow-routing algorithm as flow can only transport from one cell to one other cell. However, multiple flow algorithms also exist that can transport flow from a cell to two or more other cells.

Tarboton (1997) reviewed different flow routing algorithms stating that the D8 algorithm has disadvantages due to discretization of flow into one of only eight directions separated by 45°. The ρ8 algorithm (Fairfield and Laymarie, 1991) is also a single flow-routing algorithm that attempts to overcome the straight path problem whilst maintaining simplicity. Flow is randomly assigned from the centre grid cell to one of its downslope neighbours with the probability proportional to the slope. However, the lack of a deterministic result is a drawback for most applications.

Multiple flow-routing algorithms allocate flow fractionally to each lower neighbour in proportion to the slope but have the disadvantage that flow is dispersed to every cell lower than the focus cell. Examples include MFD (Freeman, 1991) that partitions flow to all down-slope neighbours by a function of slope to an exponent. The DEMON algorithm (Costa-Cabral and Burges, 1994) is an extension of a method by Lea (1992) where grid elevation values are taken as pixel corners and flow directions based on the aspect of a plane surface fit to each pixel. Tarboton's (1997) own D∞ (or DInfinity) algorithm routes flow in the direction of steepest descent of the eight triangular facets formed in 3*3 grid cell window.

In general, multiple flow algorithms give better performance than single flow algorithms in hydrological studies (Erskine et al., 2006; Erskine et al., 2007) but have some disadvantages over D8. These include inefficiency of data storage and robustness including the inability to cope with difficult data areas e.g. a saddle (Tarboton, 1997). D8 performs well in minimising dispersion, simple and efficient grid storage and robustness, but introduces grid bias and flow directions are not precisely resolved. However, McMaster (2002) concluded that D8 and D∞ methodologies predicted stream locations as accurately as each other. For these reasons, although alternatives exist and there are limitations of D8, the D8 routing algorithm is considered acceptable.

3.4.3 Impact of grid resolution on flow routing

Resolution of a grid-based DEM impacts on the ability to accurately direct flow across the model. Horrit and Bates (2001) compared water levels and hydrometric response of an inundated area between those observed and those modelled using various model sizes of DEM. Consistent performance up to 500m grid size DEM was witnessed with a significant drop in performance with grid cell size above 500m.

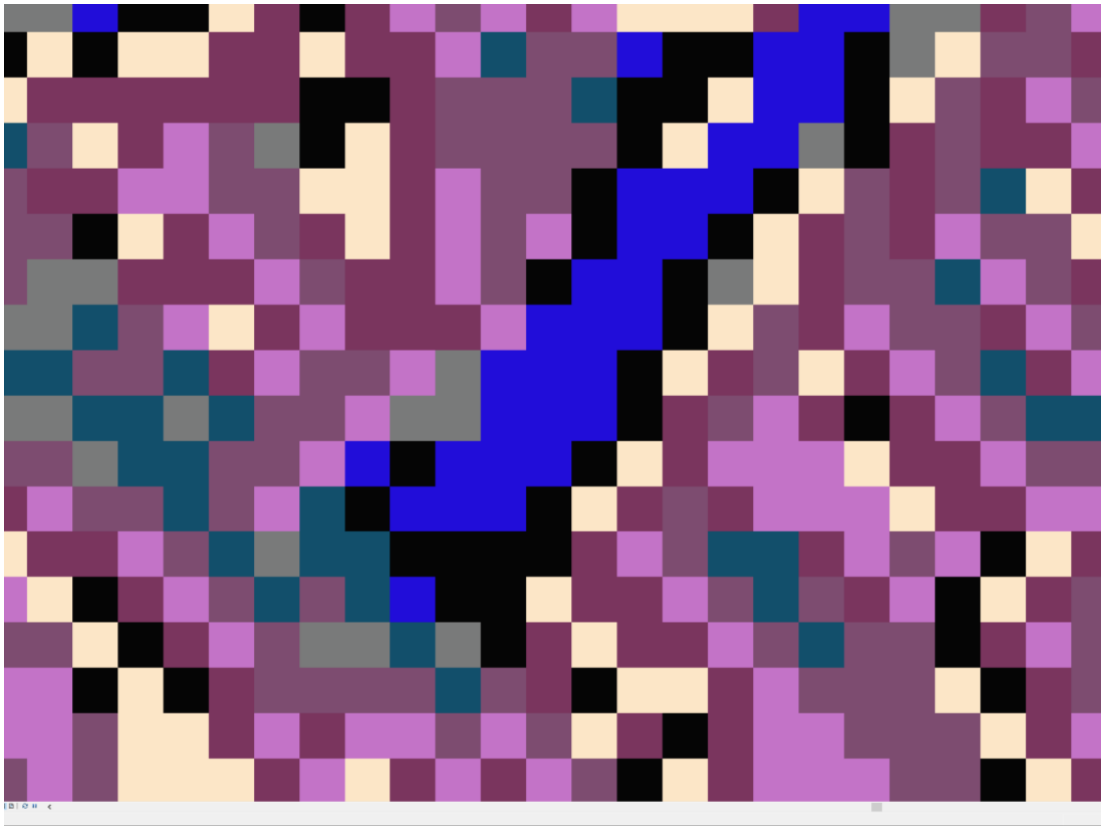
McMaster (2002) compared average distance between predicted and mapped stream channels with an optimum found at 180m resolution DEM, with little improvement as resolution is increased further. Above 180m resolution the distances began to increase. Although the 15-arcsecond resolution DEM of the HydroSHEDS suite is approximately 450m, the river network has been derived from a 3-arcsecond DEM (~90m) and hence should not suffer too many of the limitations of decreased resolution.

With all grid-based routing algorithms, a cell is considered river or not river depending on catchment area or calculated flow. As coarser resolutions are employed it becomes increasingly difficult to distinguish between land and river as river profiles are ignored. Hence a cell may be designated as river but the river itself may only be a few metres wide.

3.4.4 Selection of a Flow Direction Raster

As stated earlier, an FDR can be generated directly from a DEM. The ArcHydro package within ArcGIS (for example) has a 'Flow Direction' tool which generates an FDR from a hydrological DEM. Although this is relatively fast, using the selected DEM (HDEM15s) would generate a poor-quality FDR as this was upscaled from a DEM at 3s using the techniques discussed in section 3.4. Therefore, a more appropriate choice would be the Asia 15s GRID FDR available directly from HydroSHEDS (referenced as HFDR15s). The values within the raster use the ESRI convention of 1 to 128 (see 3.4.1) with endorheic basins labelled as -1.

Zooming the image on a region of China (see figure 3.11) shows how different colours represent the flow direction of each cell (note that the colours used to represent the flow directions can be modified by the user).



■ = 1 (East), ■ = 2 (South East), ■ = 4 (South), ■ = 8 (South West)

■ = 16 (West), ■ = 32 (North West), ■ = 64 (North), ■ = 128 (North East)

Figure 3.11: HydroSHEDS 15-arcsecond resolution Flow Direction Raster Asia (HFDR15s) as reproduced within ArcGIS and magnified to a portion of China

3.4.5 Flow Accumulation

In the D8 model, water flows from cell to cell in the direction of the FDR and accumulates as water moves down the catchment towards the catchment outlet. At the catchment boundaries (or watershed) there is only accumulation from one cell, whereas at the catchment outlet there will be accumulation of flow from many cells, and in the case of a river such as the Yangtze, millions of cells. Flow accumulation can be used to depict the river network assuming rivers form when there is sufficient upslope catchment, or sufficient accumulation of cells.

To demonstrate this methodology, the example flow direction grid from figure 3.10 will be used (see figure 3.12 a). Figure 3.12 b represents the accumulation of flow with an example river threshold of 3 cells accumulation (highlighted in blue), and shows the dramatic increase

in accumulation as tributaries meet to form larger rivers. The river system becomes clearer if the cells are converted to a polyline joining river cells in the direction of flow, as in c).

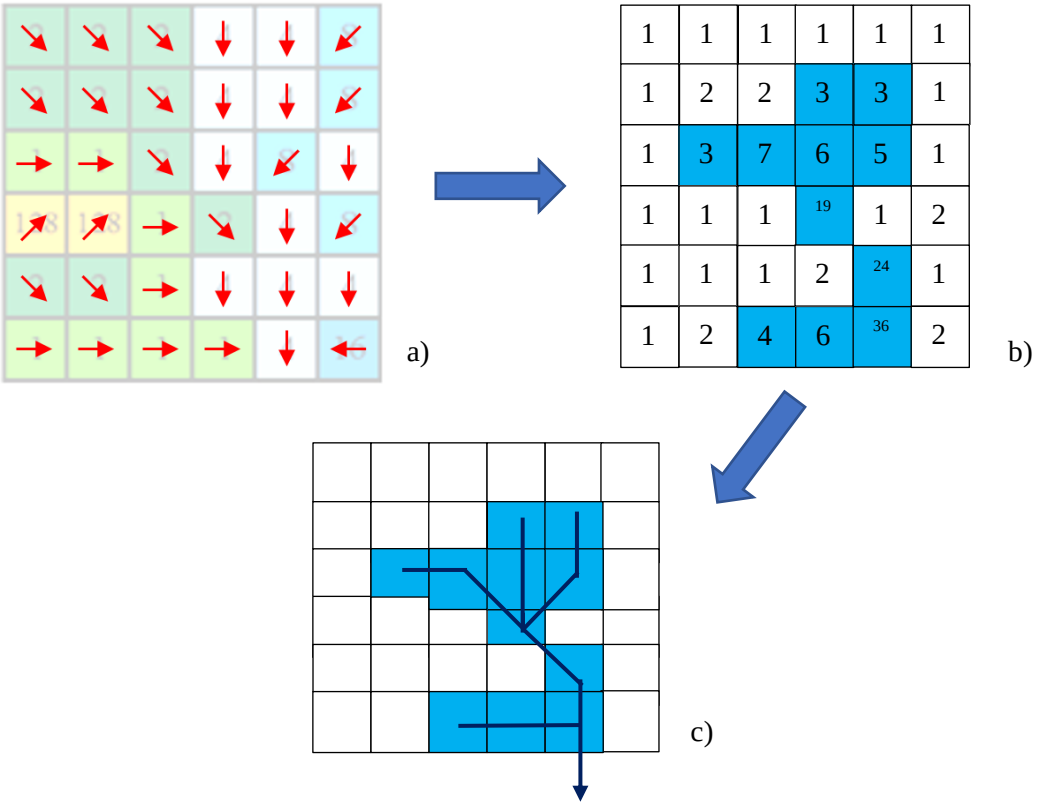


Figure 3.12: Example of methodology of deriving river networks from a flow direction grid (a) and flow accumulation (b) with river cells designated with a threshold of 3 highlighted in blue. River pathways are easier to see if the cells are converted to a polyline joining the river cells in the direction of flow (c).

In the HydroSHEDS data suite the threshold of river is set to 8km^2 which is too large for this purpose, hence a new river dataset with a catchment of 2km^2 was created employing tools within ArcHydro but still using the HydroSHEDS DEM and FDR data.

The HFDR15s raster was used to derive a 15-arcsecond resolution flow accumulation raster (FAR) (referenced as FAR15s) using the 'Flow Accumulation' tool within ArcHydro (which took several hours to compute). Due to the logarithmic increase in flow accumulation in rivers, particularly at the catchment outlet, when plotted in ArcGIS with a colour scheme stretching from black (minimum accumulation) to white (maximum accumulation), the image appears mostly black as most cells have a relatively low accumulation compared to the

catchment outlet. On zooming the image in the region of the catchment outlet, the main Yangtze River is visible as white due to the large accumulation of flow (see figure 3.13).

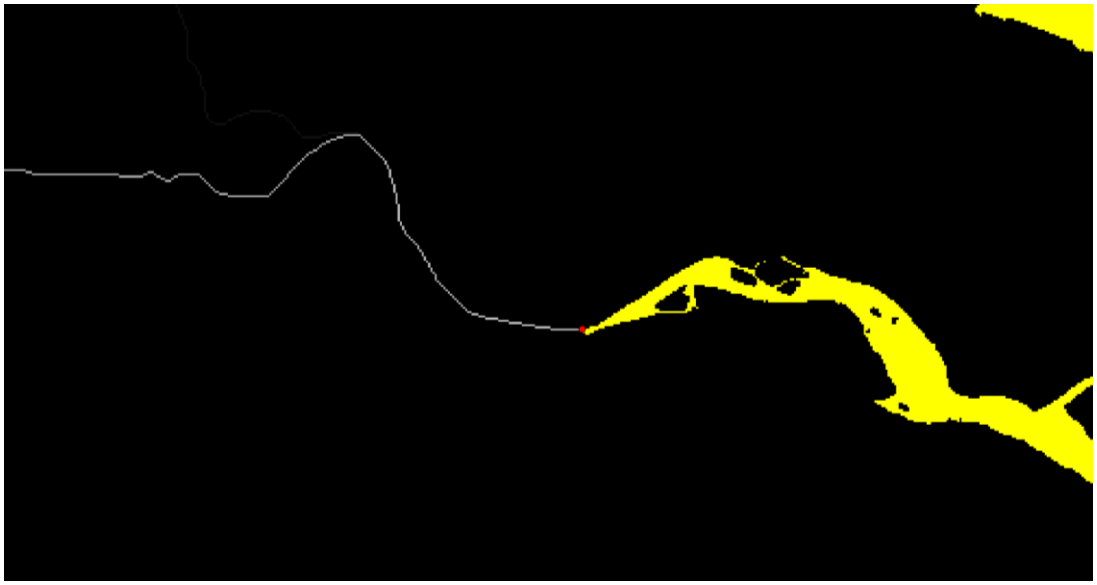


Figure 3.13: Flow Accumulation Raster (FAR) at the Yangtze mouth as derived from the HydroSHEDS 15-arcseconds Flow Direction Raster (HFDR15s) as reproduced within ArcGIS showing the main Yangtze River in white and no data areas (i.e. sea) in yellow

Interrogation of the FAR15s shows flow accumulation at the mouth of the Yangtze is equal to 10,320,449 cells. At the mouth of the Yangtze, 15-arcseonds in longitude is approximately 0.4404km and 15-arcseconds in latitude is approximately 0.4633km calculated by the Harversine formula (Moveable Type Scripts, 2016). Hence a cell area at 15-arcseconds resolution is approximately 0.204km^2 (see figure 3.14).

The Harversine formula calculates the great-circle distance between two points i.e. the shortest distance over the Earth's surface and the formula is as follows:

$$\text{Harversine formula: } a = sin^2\left(\frac{\Delta\varphi}{2}\right) + cos\varphi_1 . cos\varphi_2 . sin^2\left(\frac{\Delta\lambda}{2}\right)$$

$$c = 2.atan^2(\sqrt{a}, \sqrt{1-a}) \qquad (3.3)$$

$$d = R.c$$

where φ is latitude, λ is longitude, with latitude and longitude values in radians, R is the Earth's radius [mean = 6371km], d is distance [km].

Hence the catchment area of the YDB as computed by the FAR15s is:

YDB Area = 0.204km^2 x 10,320,449 = 2,105,372km^2

Note this is only an approximation as cell area varies with latitude and cells becoming increasingly smaller as they move away from the equator.

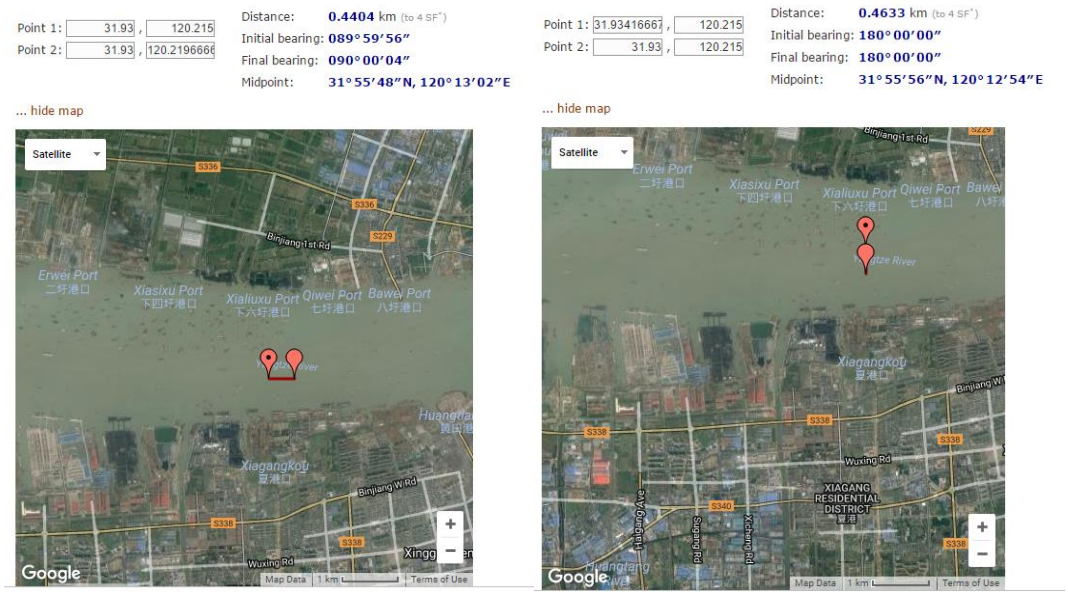


Figure 3.14: 15-arcseconds difference in longitude (left) and latitude (right) at the mouth of the Yangtze River and measurements calculated by the Harversine formula (Moveable Type Scripts, 2016)

3.4.6 Defining the Yangtze River drainage basin

Up to now, all computations and datasets have been made at the Asia scale, analysing far more data than necessary and hence it would be desirable to do all modelling at the size of the area of interest i.e. the YDB. Now that flow accumulation has been established, it is possible to derive the contributing area (i.e. the drainage basin area) to the Yangtze mouth, or pourpoint (the point in which all flow pours).

The YDB can be derived using the ArcGIS Spatial Analysist 'Watershed' tool. A pourpoint was manually created and placed on the cell of maximum flow accumulation (see red dot in figure 3.13), easily identified due to having the maximum accumulation and entering the estuary in the region expected.

Inputs to the 'Watershed' tool include the HFDR15s and the pourpoint created, resulting in a raster dataset of the YDB. Figure 3.15 shows the basin overlaid on top of the original Asia HDEM15s to show its location and size relative to Asia. This will be referred to as the YDB15s (Yangtze drainage basin at 15-arcsecond resolution).

Closer examination of the YDB15s raster produced shows no-data areas within the drainage basin resulting from endorheic basins (a closed drainage basin that normally retains water and allows no outflow to other external bodies of water). As these are small in area relative to the total area, no further exploration of these was made. This raster was converted to a polygon shape (using the Raster to Polygon tool) which can be used to crop other datasets to the YDB.

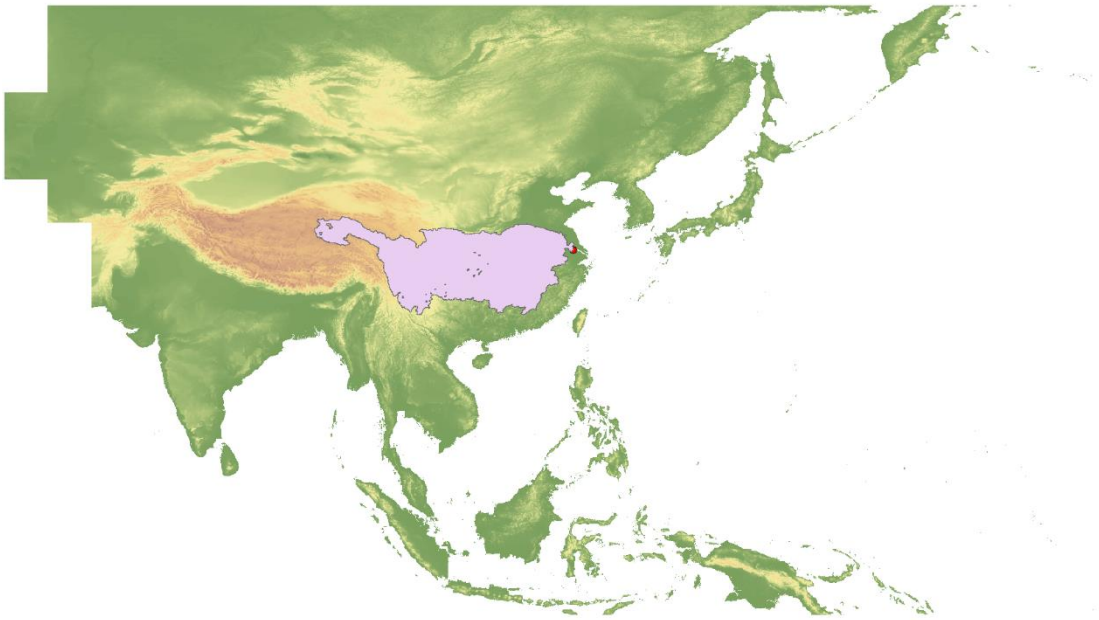


Figure 3.15: Derived Yangtze drainage basin area raster using the ArcGIS Spatial Analyst Watershed tool at 15-arcseconds resolution (YDB15s) superimposed on HydroSHEDS DEM (HDEM15s)

3.4.7 Cropping datasets to the Yangtze drainage basin area

Cropping the previously derived datasets to that of the YDB is achieved using ArcGIS' Clip (Analysis) tool which extracts input features that overlay the clip features. Therefore, the YDB15s polygon can be used as a 'biscuit-cutter' to derive a Yangtze DEM, Yangtze FDR and Yangtze FAR from the original HDEM15s, HFDR15s and FAR15s rasters. These new datasets wills be referred to as YDEM15s, YFDR15s and YFAR15s.

The YDEM15s figures (see Figure 3.16, 3.17) show high mountains in the western third of the basin with a maximum elevation of 6482m in Qinghai close to the Tibetan border. When overlaid on a provincial map of China (Figure 3.18) the basin includes part of southern Qinghai, a small section of the far east of Tibet, most of Sichuan, the northern part of Yunnan, the northern half of Guizhou, all of Chongqing, part of southern Gansu and southern Shaanxi,

all of Hubei and Hunan, a large proportion of Henan, Jiangxi and Anhui and part of western Jiangsu. It is estimated that the basin is home to approximately 1/3 of the population of China.

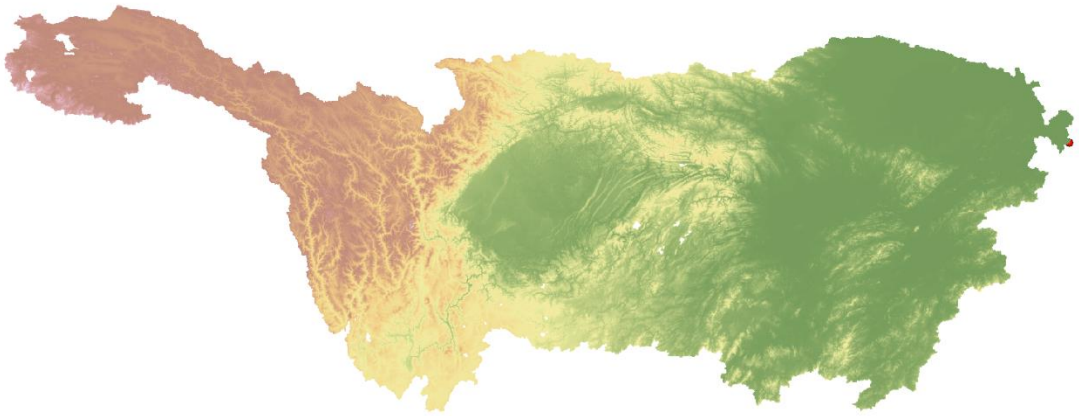
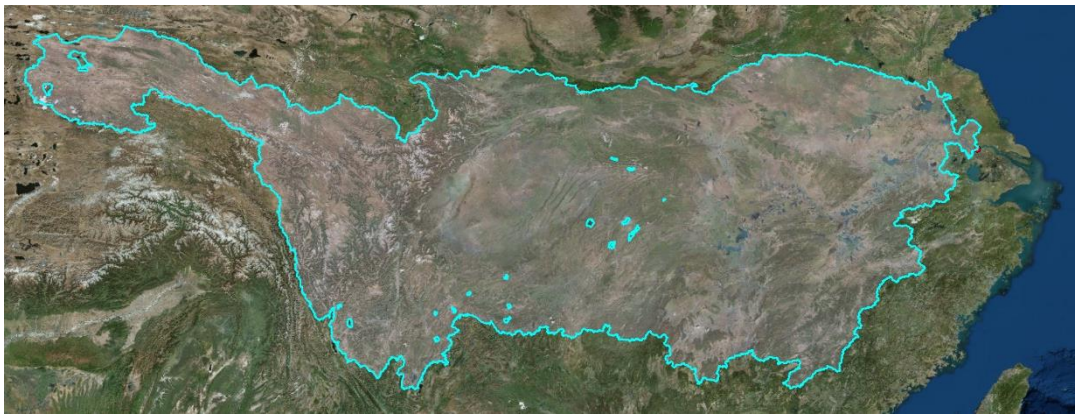


Figure 3.16: Yangtze drainage basin DEM (YDEM15s) clipped using the ArcGIS Clip (Analysis) tool at 15-arcseconds resolution with highland regions in red/brown and lowland area in green



Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community

Figure 3.17: Yangtze drainage basin DEM (YDEM15s) projected onto ArcGIS basemap satellite imagery

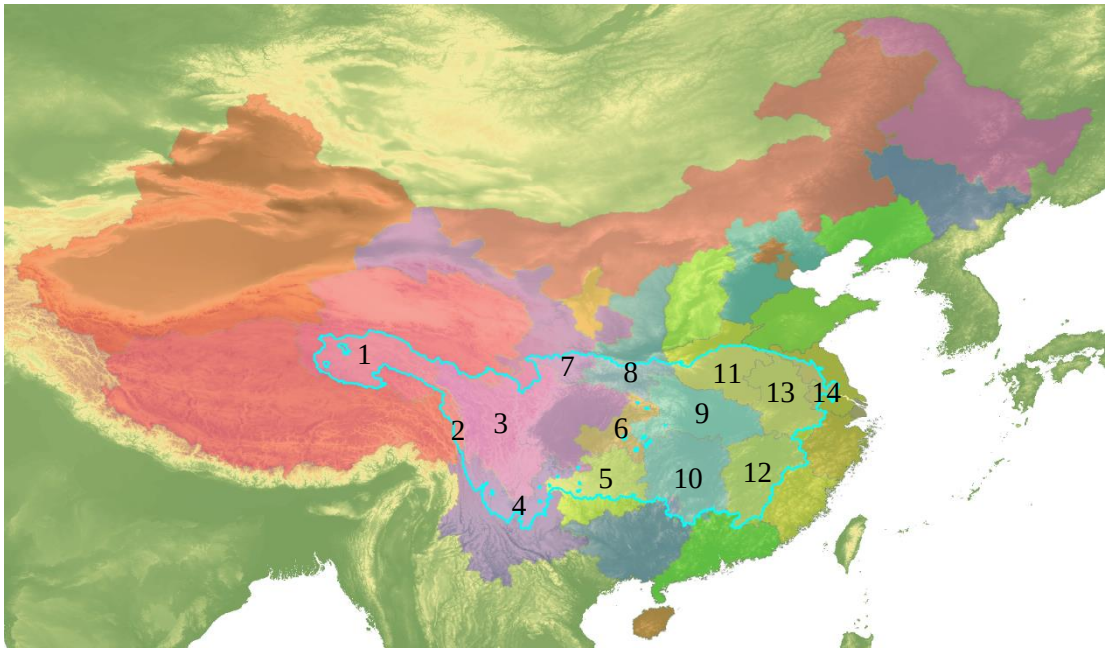


Figure 3.18: Yangtze drainage basin DEM (YDEM15s) projected onto a China provinces map

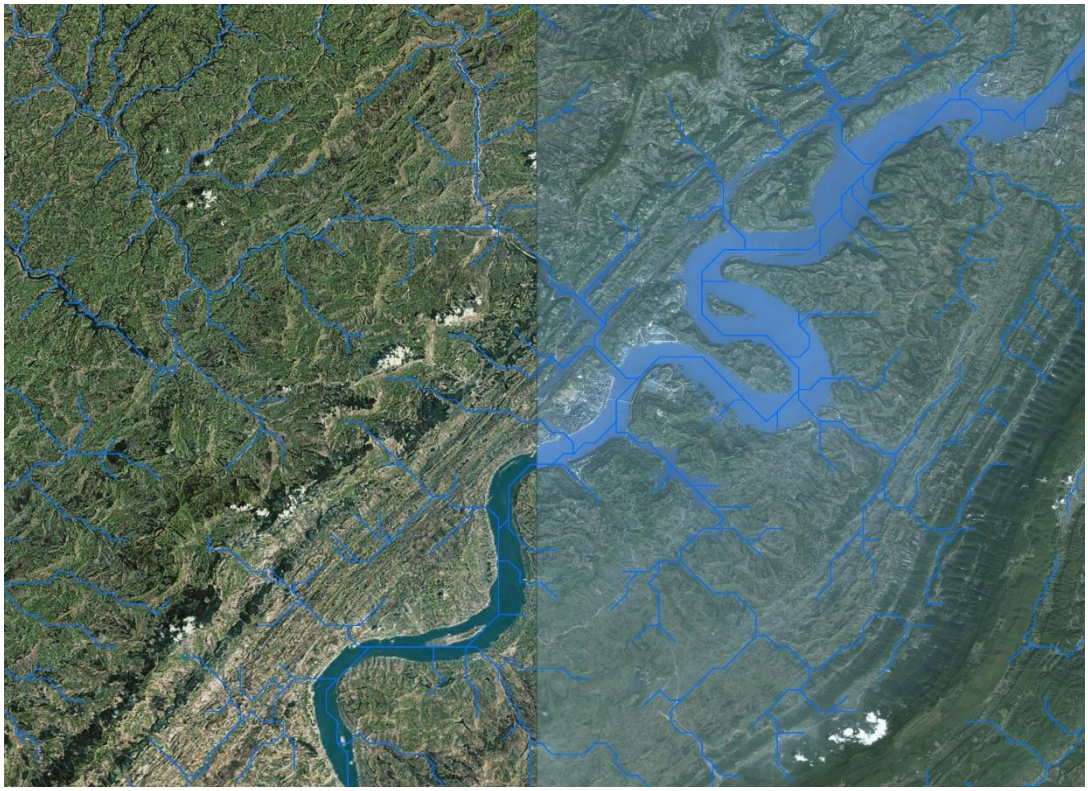
(1 = Qinghai, 2 = Tibet, 3 = Sichuan, 4 = Yunnan, 5 = Guizhou,
6 = Chongqing, 7 = Gansu, 8 = Shaanxi, 9 = Hubei, 10 = Hunan, 11 = Henan,
12 = Jiangxi, 13 = Anhui, 14 = Jiangsu)

3.4.8 Deriving the Yangtze drainage basin river dataset

Using the 'Stream definition' function of ArcHydro with YFAR15s dataset as input, a river network can be determined with a specified threshold accumulation, producing a raster image of the river network. A 2km^2 catchment threshold requires a minimum of 10 cells accumulation and, for clarity, this was converted to a polyline vector image using the 'Raster to Polyline' tool, resulting in a dense river network.

Some of these 2km^2 catchment rivers may not be perennial and may only flow when precipitation is heavy and high levels of surface runoff generated. However, as flow conditions are yet unknown, it was decided to keep this threshold to ensure river sites are not missed in the hydropower search. It is possible that rivers form from smaller upslope catchment areas, but it is unlikely these would be suitable for economically viable hydropower. Zooming in on a region and superimposing the river network onto a satellite map shows that many of the river polylines generated do follow river courses on the satellite

image, although smaller reaches appear over land, hence unlikely to be perennial (see figure 3.19).



Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community

Figure 3.19: Yangtze drainage basin (YDB) river network 15-arcsecond polyline vector image with a threshold of 2km^2 catchment (10 cells at 15-arcseconds) superimposed on ArcGIS World Imagery basemap graphics zoomed in a region around 30.198 latitude and 108.161 longitude

A larger threshold can be set to view major rivers of the Yangtze, as shown in figure 3.20, with blue lines representing a threshold of 10,000 cells (approximately 2,000km^2) and red lines 100,000 cells (approximately 20,000km^2) and superimposed on a map of China. The main Yangtze River (highlighted) starts high on the Qinghai plateau, flows south along the Tibet/Yunnan border towards Lijiang, north-easterly through Chongqing, east through Yichang and Wuhan, north-east towards Nanjing, entering the estuary between Nanjing and Shanghai.


QINGHAI
Xining
Lanzhou
SHAANXI
SHANXI
HENAN
SHANDONG
JIANGSU
ANHUI
HUBEI
HUNAN
JIANGXI
FUJIAN
ZHEJIANG
C H I N A
SICHUAN
YUNNAN
Chengdu
Nanchong
Wuhan
Shanghai
Hangzhou
Ningbo
Changsha
Nanchang
Fuzhou
Kunming
Guilin
Liuzhou
GUANGXI
GUANGDONG
Taipei
Taichung
Xiamen
Jieyang
Wenzhou
Pingxiang
Liuyang
Lichuan
Yueyang
Zhengzhou
Kaifeng
Luoyang
Xi'an
Xianyang
Zhumadian
Xinyang
Hefei
Nanjing
Wuxi
Suzhou
Yangzhou
Taizhou
Yancheng
Huaiyin
Linyi
Jining
Zaozhuang
Jinan
Zibo
Weifang
Qingdao
Tai'an
Handan
Changzhi
Anyang
Puyang
Xinxiang
Jiaozuo
Xuzhou
Huaibei
Shaoxing
Chuzhou
Huainan
Fengyang
Shaoyang
Zhuzhou
Yongkang
Quanzhou
Taoyuan
Jin-men
Lhasa
Wangda
Zigong
Meijiang
Qijiang
Liupanshui
Liuzhi
Liuyang
Shaoguan
Guiyang
Liuzhi
Liupanshui
Liuyang
Pingxiang
Lichuan
Yongkang
Quanzhou
Taoyuan
Jin-men
Taichung
Xiamen
Jieyang
Wenzhou
Fuzhou
Ningbo
Shaoxing
Hangzhou
Changsha
Nanchang
Yueyang
Zhengzhou
Kaifeng
Luoyang
Xi'an
Xianyang
Zhumadian
Xinyang
Hefei
Nanjing
Wuxi
Suzhou
Yangzhou
Taizhou
Yancheng
Huaiyin
Linyi
Jining
Zaozhuang
Jinan
Zibo
Weifang
Qingdao
Tai'an
Handan
Changzhi
Anyang
Puyang
Xinxiang
Jiaozuo
Xuzhou
Huaibei
Shaoxing
Chuzhou
Huainan
Fengyang
Shaoyang
Zhuzhou

River name	Gauging Station	Published Catchment Area (km^2)	HydroSHEDS Catchment Area (km^2)	Error difference %
Bailu He	Baiqueyua	284	290	2.4
Lianshui He	Yangxinjiang	535	695	29.9
Zagunao He	Zagunao	2,404	2,400	-0.2
Huai He	Changtaigua	3,090	3,025	-2.1
Laoguan He	Xixia	3,418	3,216	-5.9
Yalong Jiang	Luning	58,943	107,882	83
Huai He	Bengbu	121,330	122,436	0.9
Tongtian He	Zhimenda	137,704	134,399	-2.4
Yangtze River	Datong	1,705,383	1,679,569	-1.5

Table 3.2: Comparison of HydroSHEDS calculated sub catchment areas within the Yangtze drainage basin (YDB) with published gauging station areas from the GRDC (GRDC, 2017)

Most of the results show good agreement with error below 3%. Two exceptions are the Lianshui He Yangxinjiang and Yalong Jiang Luning catchments where it is most likely that the published data is erroneous as the latitude/longitude coordinates given do not line up with a river. GRDC could not offer an explanation as they publish the data as given.

Another method of validating the dataset, and the underlying DEM for spot height retrieval, is to compare the river network to satellite imagery. River confluences are easy to identify both on the derived river network and on satellite imagery and the locations and heights of river confluences were compared between the satellite image and the derived network.

The HydroSHEDS derived river network of 10,000 cells (2,000km^2 catchment – see section 3.4.8) shapefile was exported into Google Earth (see figure 3.21) as Google Earth is much faster at rendering satellite imagery than using ArcGIS basemap imagery and includes an in-built DEM. The 10,000-cell network was selected as these rivers are big enough to easily identify on the satellite image. Google Earth base imagery is 30m multispectral Landsat (USGS, 2016) pan-sharpened with 15m Landsat imagery. Google is in the process of upgrading this imagery with 2.5m SPOTImage imagery (Airbus, 2016). In some areas, it is even higher resolution to less than a metre.

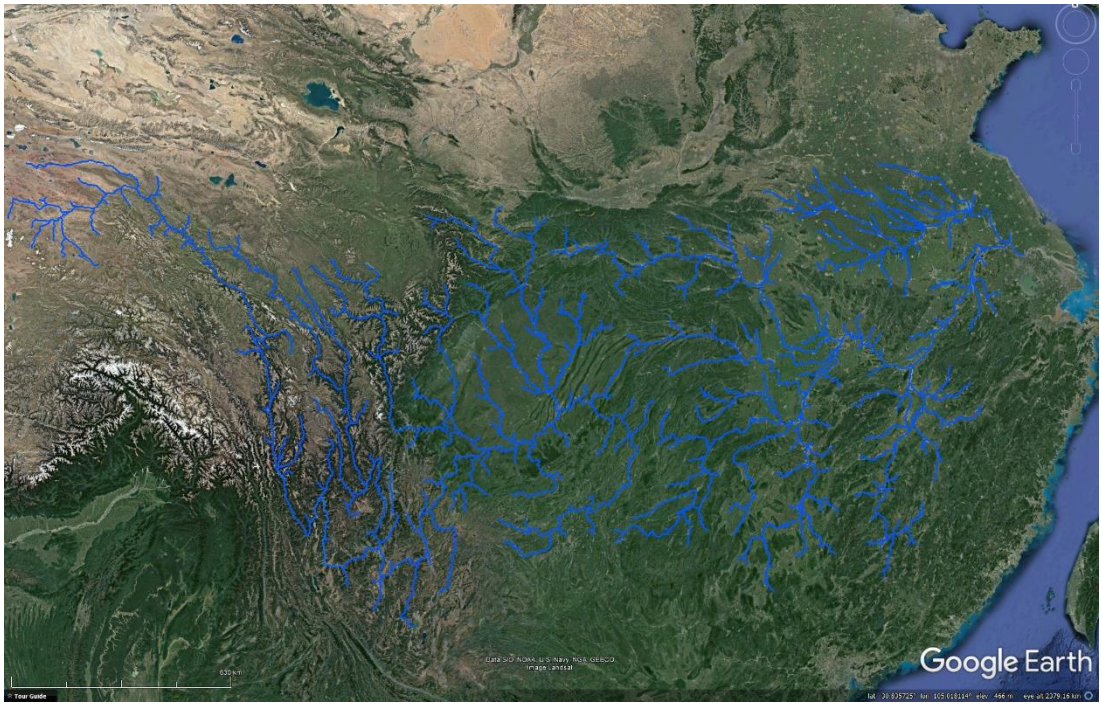


Figure 3.21: Derived river network at 10,000 cells (~2,000km^2 catchment) exported into Google Earth (Google, 2016)

Starting at the far west of the river network, each confluence along the main stream was examined and the longitude/latitude of both the HydroSHEDS derived confluence and the satellite image confluence (if clear enough) were extracted. The confluence point of the satellite image was estimated as being the midpoint of where both streams meet (see figure 3.22). Using the ruler tool, the distance of the two points was then measured. If the derived confluence was very close to the image confluence it was given a value of 0m. The Google Earth reported height of the image confluence was also obtained and compared with the height of the derived confluence using the HDEM15s data within ArcGIS. This was repeated for all confluences on the main Yangtze stream until the lower reaches. Table 3.3 (overleaf) shows the distances and DEM error of all the confluences measured.

The distance between the HydroSHEDS derived river network and the satellite image is less than 500m in 74% of the confluences analysed. As one cell width is nearly 500m, the resolution of the underlying dataset is too coarse to perform better. 11% of the derived confluences analysed are within 1km of the satellite confluence i.e. within two cells distance. The remaining 15% are beyond 1km distance with one site 11km from the actual confluence. A common problem is when tributaries run parallel to the main stream (as in No. 23) with a

small ridge of land separating the two rivers. In the HydroSHEDS averaging process the joining river breaks through too early showing a distance discrepancy between the predicted confluence and the actual confluence (see figure 3.23).

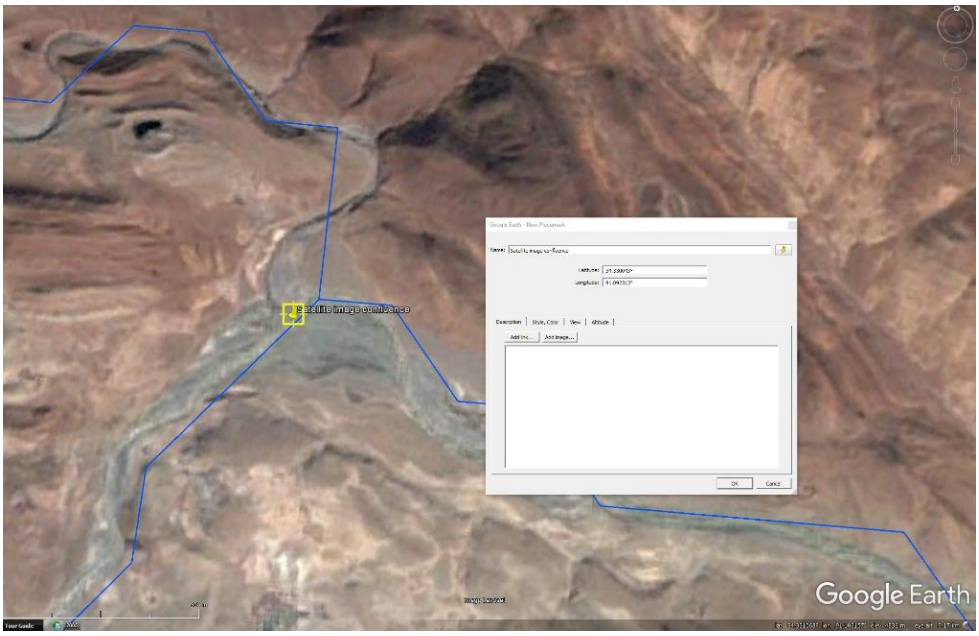


Figure 3.22: Extracting latitude and longitude data of a confluence on the Google Earth satellite image

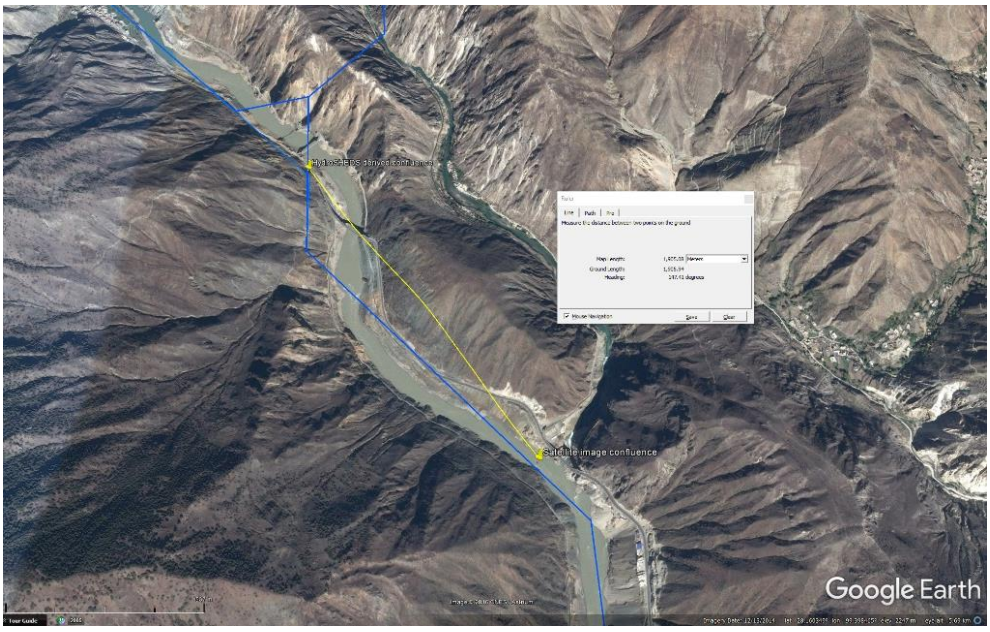


Figure 3.23: Error between Google Earth satellite image confluence (No.23) and the HydroSHEDS derived river network confluence due to thin strips of land between tributary and the main stream

No.	Confluence Latitude Google Earth	Confluence Longitude Google Earth	Confluence Latitude HydroSHEDS	Confluence Longitude HydroSHEDS	Distance between Google Earth and HydroSHEDS	Confluence Google Earth Height	Confluence HydroSHEDS Height	Height Error
1	34.331195°	91.093736°	34.331195°	91.093736°	0	4786	4778	-8
2	34.328122°	92.116008°	34.339575°	92.106263°	1600	4598	4595	-3
3	34.094098°	92.913890°	34.089561°	92.918639°	500	4477	4472	-5
4	34.201925°	93.701956°	34.201925°	93.701956°	0	4393	4390	-3
5	34.395778°	93.906777°	34.393773°	93.910376°	267	4366	4360	-6
6	34.555392°	94.055556°	34.547880°	94.047727°	1158	4334	4330	-4
7	34.681018°	94.460253°	34.681018°	94.460253°	0	4288	4286	-2
8	34.669162°	94.931219°	34.668741°	94.931153°	63	4221	4219	-2
9	34.481237°	95.348472°	34.481175°	95.352051°	305	4165	4173	+8
10	34.023776°	95.828431°	34.022900°	95.822846°	480	4063	4060	-3
11	33.803215°	96.049423°	33.806194°	96.047954°	361	4009	4006	-3
12	33.723222°	96.406236°	33.731252°	96.410417°	981	3875	3945	+70
13	33.435091°	96.718097°	33.430900°	96.702153°	1533	3731	3756	+25
14	32.978490°	97.243626°	32.977081°	97.243732°	160	3530	3555	+25
15	31.380129°	98.884145°	31.377063°	98.885402°	291	2978	3006	+8
16	31.251216°	98.776839°	31.256092°	98.777084°	570	2952	2953	+1
17	31.193966°	98.600198°	31.193776°	98.602075°	135	2907	2927	+20
18	31.035406°	98.760400°	31.035395°	98.764526°	396	2875	2907	+32
19	29.937225°	99.058336°	29.939484°	99.056218°	320	2495	2514	+19
20	29.758040°	99.007532°	29.760522°	99.006226°	283	2487	2482	-5
21	28.982911°	99.123946°	28.985277°	99.122900°	305	2269	2295	+26
22	28.353576°	99.229325°	28.356249°	99.227064°	409	2045	2114	+69
23	28.154477°	99.400013°	28.172862°	99.389442°	2277	2001	2073	+62
24	27.167512°	100.067091°	27.164575°	100.068701°	380	1810	1877	+67
25	27.757590°	100.375331°	27.760236°	100.377154°	330	1480	1529	+49
26	26.797351°	100.435238°	26.797766°	100.439583°	437	1304	1340	+36
27	26.169965°	100.909200°	26.164568°	100.910394°	656	1127	1205	+78
28	26.726989°	101.860324°	26.726989°	101.860324°	0	996	1040	+54
29	26.606173°	101.802099°	26.605075°	101.802170°	126	986	989	+13
30	25.960512°	101.872040°	25.960154°	101.877136°	592	929	960	+31
31	26.110306°	102.134384°	26.110457°	102.131264°	362	894	905	+11
32	26.306361°	102.803071°	26.306042°	102.806202°	312	759	808	+49
33	26.544560°	103.053552°	26.526747°	103.056318°	2022	698	745	+47
34	26.779479°	102.989858°	26.781121°	102.989573°	190	660	663	+3
35	26.968620°	102.891056°	26.968682°	102.889611°	166	630	684	+54
36	27.417271°	102.952345°	27.418693°	102.952083°	193	562	654	+92
37	27.425067°	103.140609°	27.426746°	103.143647°	330	527	551	+24
38	28.052050°	103.427124°	28.052050°	103.427124°	0	427	444	+17
39	28.628116°	104.425130°	28.627041°	104.422918°	244	262	281	+19
40	28.766183°	104.676224°	28.764611°	104.677006°	220	252	254	+2
41	28.736578°	105.075826°	28.743689°	105.064588°	1307	237	241	+4
42	28.781193°	105.368909°	28.781193°	105.368909°	0	219	219	0
43	28.905814°	105.452098°	28.906147°	105.456178°	330	219	220	+1
44	28.804816°	105.845980°	28.806167°	105.843790°	253	197	205	+8
45	29.289065°	106.389157°	29.289065°	106.389157°	0	169	172	+3

No.	Confluence Latitude Google Earth	Confluence Longitude Google Earth	Confluence Latitude HydroSHEDS	Confluence Longitude HydroSHEDS	Distance between Google Earth and HydroSHEDS	Confluence Google Earth Height	Confluence HydroSHEDS Height	Height Error
46	29.573226°	106.583856°	29.573226°	106.583856°	500	155	181	+26
47	29.659699°	106.890494°	29.606010°	106.785704°	11000	154	163	+9
48	29.810243°	107.081229°	29.810243°	107.081229°	0	145	144	-1
49	29.718278°	107.399818°	29.722592°	107.402032°	615	136	135	-1
50	29.890580°	107.737468°	29.897553°	107.739368°	700	130	130	0
51	30.950286°	108.656919°	30.943502°	108.668743°	1144	98	97	-1
52	29.457687°	113.144642°	29.457687°	113.144642°	0	19	18	-1
53	30.566577°	114.288708°	30.564568°	114.285396°	371	16	17	+1

Table 3.3: Distance between Google Earth satellite image confluences on the Yangtze main stream and HydroSHEDs derived confluences and height error – large distance error highlighted

In the case of No.47 (see figure 3.24), the HydroSHEDS model shows the tributary heading south into the main stream whereas in the satellite image the river turns east and joins several km further down the main stream. It is not clear whether this is due to errors in the development of the HydroSHEDS data or whether the tributary is in some way diverted by man-made structures. This is the only gross error in the confluence points examined and overall there is a good fit between the HydroSHEDS derived river network and satellite imagery. If the No.47 outlier is removed from the data, the mean error distance is 464m, approximately 1 cell width.

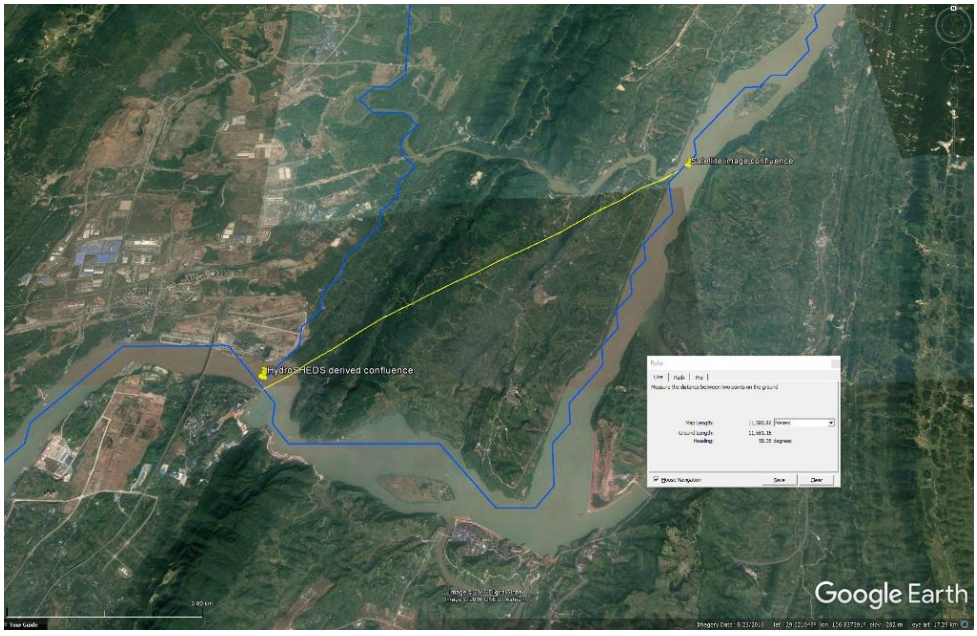


Figure 3.24: Error between Google Earth satellite image confluence (No.47) and the HydroSHEDS derived river network confluence due to routing the tributary erroneously south

Absolute mean error between the HydroSHEDS derived confluences DEM elevation and the Google Earth satellite image confluence elevation is 21.2m, with most values higher than the Google Earth reported elevation (although it is important to note that Google Earth is not error free). The fact that the HydroSHEDS derived data is higher comes as no surprise. Rivers form at the lowest points of an area and if the resolution of the DEM dataset is decreased, then the average height of a cell will naturally increase compared to the river due to taking in more of the valley sides. Note how in the upper regions (>4000m) the error is much smaller (mean of 4.2m) and similarly in the lower regions (<250m) (mean of 3.7m). Upper regions of the YDB are characterised by plateau and hence when upscaling the HydroSHEDS data there is less inclusion of steep valley sides to increase the height value compared to the river. Similarly, in the lower portions of the YDB, rivers run across plains and once again the averaging effect does not dramatically increase the height of the cell. Mid-regions are characterised by narrow steep mountain valleys (see figure 3.25) increasing averaged cell heights dramatically compared to the river height.

Figure 3.25: Mid-Yangtze the rivers flow through steep mountain valleys resulting in increased DEM error between Google Earth rivers and HydroSHEDS derived DEM values

Erroneous DEM values potentially impact on the estimated net head of a scheme and the air temperature influencing snowfall/melt/frost. Error in the region of <100m is unlikely to have any real impact on snowfall/melt/frost but it could inflate (or reduce) the net head, and hence inflate (or reduce) the estimated energy output from a scheme.

Overall, the difference between the HydoSHEDS derived river network and Google Earth elevations appears to be acceptable but with the caveat that potential scheme height differences should be quality checked against a high-quality DEM such as SRTM/HydroSHEDS 3-arcseconds or Google Earth (see results chapter 7; although at this stage the HydroSHEDS 15-arcseconds DEM data was considered acceptable, it was found that it grossly inflated the heights of some schemes and a novel methodology incorporating HydroSHEDS 3-arcseconds data was found to correct this, detailed in Chapter 7).

3.5 Creating a Yangtze lakes and reservoirs dataset

Knowledge of the water bodies within YDB is necessary both for costing of hydropower sites and to ensure hydropower sites are not situated within lakes. Water bodies may affect hydropower costs including routing of roads and power lines. There are several free GIS datasets available including Natural Earth lakes and reservoirs vector files at a scale of 1:10million (Natural Earth, 2016) and the GLWD mentioned in 3.3.4.1. The latter has level 1 data consisting of lakes with surface area greater than 50km^2 and reservoirs greater than 0.5km^3 and level 2 data consisting of open water bodies with surface area greater than 0.1km^2 excluding the larger lakes and reservoirs in level 1.

The GLWD level 2 data was filtered by the author to remove rivers (as rivers are of course of key interest) and a Yangtze lakes and reservoirs dataset was created by merging the three datasets into one and cropping to the YDB (see figure 3.26). A raster file of the dataset was created with values of 1 referring to a lake or reservoir.

Figure 3.26: Final Yangtze basin lakes and reservoirs dataset

3.6 Determining river flow across the Yangtze drainage basin

A key component of studying hydropower viability is an understanding of the river flow regime. Although flow data in the public domain is available for a few river sites within the YDB (for example from GRDC), there are no datasets covering all rivers at sufficient temporal and spatial resolution for the purpose required here.

GRDC offers a Global Composite Runoff Fields product (Fekete et al., 2002) which gives gridded mean annual runoff (mm/year) at a spatial resolution of 30-arcminutes. Meijer (2012) used this dataset, in conjunction with the HydroSHEDS 3-arcsecond DEM, to evaluate the global hydropower potential by finding the net head across each 3-arcsecond cell (by finding the slope of each cell) and multiplying by the 'runoff weighted' accumulation. Such a methodology is very coarse as true hydropower evaluations cannot use average flow values since hydropower sites are configured (and costed) based on an understanding of how flow varies. Rivers rarely (if ever) run near to the average flow all year-round as flow varies depending on meteorological conditions.

Hence it is necessary to generate flow duration curves (FDC) for each river cell across the basin by developing, calibrating and validating a hydrological model. An FDC is a plot that shows the percentage of time flow is likely to exceed or equal a value (see figure 3.27) and usually performed over many (20+) years to ensure the data does not represent a particularly wet or dry year. The term Q_x is used to represent the flow (Q) that is exceeded by x % of the

total time of the record. Hence Q_{100} is exceeded 100% of the time i.e. the lowest flow of the river, and Q_0 is exceeded 0% of the time i.e. maximum flood conditions.

Developing such a model to produce daily average flow values across the river network on a scale the size of the YDB is a major undertaking and hence will be discussed in subsequent chapters. Chapter 4 will discuss the meteorological dataset development. Water inputs into such a model are solely from precipitation (rain/snowfall) and therefore it is necessary to develop a Yangtze specific precipitation dataset together with a snowfall/melt model. Snowfall/melt models use temperature as input which requires a Yangtze specific temperature dataset. Before reaching the catchment outlet, water may leave the basin through evapotranspiration, a process where water is transferred from the land to the atmosphere by evaporation from land/water surfaces and transpiration from plants. Evapotranspiration datasets do not exist on a daily scale covering the Yangtze and hence Chapter 4 will also discuss development of a daily evapotranspiration dataset (see 4.4).

Chapter 5 will discuss the selection, implementation, calibration and validation of a hydrological model with outputs as daily flow rasters, used to establish a basin wide FDC covering all rivers.

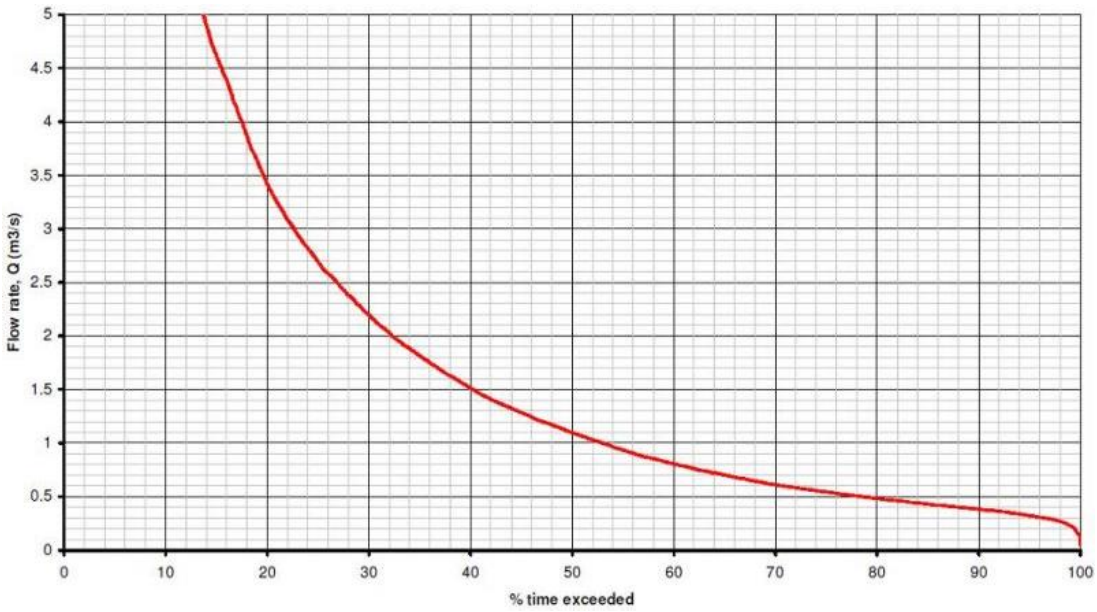


Figure 3.27: An example flow duration curve (FDC) showing flow (vertical axis) and % time exceeded (horizontal axis)

3.7 River width and depth information

The modelling techniques discussed give no information to the size of the rivers being modelled as they are all considered one cell width. Hence, when plotted, a major river such as the Yangtze will appear the same size as a small stream. River width and depth information is important in costing hydropower configurations particularly regarding civil works (i.e. weir construction). Datasets consisting of width and depth information of rivers do not exist (or at least not within the YDB). Satellite imagery is too low resolution to estimate river widths and, without an automated methodology, would take far too long to manually measure. Even if it was possible, this wouldn't give information about river depth.

Although rivers of equal flow vary in their cross-section, an estimation of the width and depth can be made using power-law relationships with discharge (Andreadis et al., 2013). It is assumed all rivers are single channel rivers and the equations given are:

$$Width \propto 7.2Q^{0.5}; Depth \propto 0.27Q^{0.3} \quad (3.4)$$

where Q is equal to flow [m^3s^{-1}]. Hence an attempt can be made to size rivers depend on their flow values. Bank full discharge is estimated to occur between Q_0 (i.e. maximum flood conditions) and Q_5, and in this project, is set to Q_5.

3.8 Development of an access road dataset

Unless at very small scale (e.g. pico hydro), most hydropower installations will require access to the site by road for installation of civil works and equipment. This does not necessarily need to be a paved road, and unpaved roads will suffice. In some instances, there may be existing roads at the site of a potential hydropower installation, but in many cases a new road will need to be laid and connected to the original road network. Testing for hydropower viability requires costings of any such roads.

A road map of China was created by the author in ArcGIS (as a shapefile) and cropped to the YDB by merging data from the following free and public domain GIS sources (see figure 3.28):

1. 2009 road shapefile within Chinamap (Center for Geographic Analysis, 2015)
2. Transportation shapefile from the GISDataDepot (Geocommunity, 2014)

3. A roads shapefile from the Harvard Faculty of Arts and Sciences (Harvard University, 2016)
4. Open Street map data downloaded via the Geofabrik downloads website (Geofabrik, 2015)
5. China roads data from the DIVA-GIS portal (DIVA-GIS, 2016).

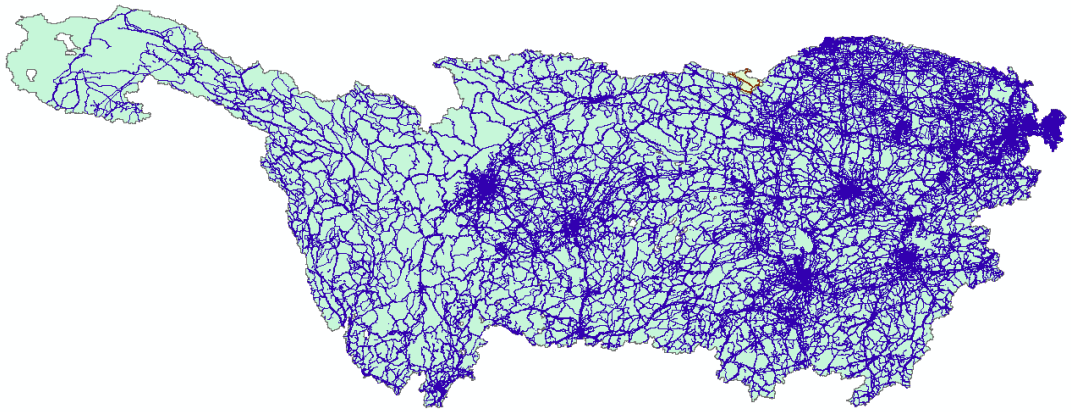


Figure 3.28: GIS road map of the Yangtze drainage basin created from numerous sources

It is possible that some unpaved roads are missing from this dataset, but no attempt was made to check for missing roads or validate the data, as there is no master reference to compare to.

3.9 Development of an electrical transmission connection dataset

A hydropower site needs to transfer the energy generated via a transmission line from the power house to the existing grid. To date it has not been possible to find a publicly available GIS dataset representing China transmission lines at all voltages. GIS datasets representing 220kV and greater transmission lines are available (e.g. Center for Geographic Analysis, 2015) (see figure 3.29) but a more thorough network representing lower voltage transmission is required for accurate costings.

Figure 3.29: China 220kV and up transmission line GIS dataset (Center for Geographic Analysis, 2015)

A shapefile of the 200kV and greater transmission network of China is available from DIVA-GIS (DIVA-GIS$_a$, 2016), and within the YDB 220kV (AC), 330kV (AC), 500kV (AC) and 500kV± (DC) transmission lines are represented. This has been filtered by the author to produce three datasets:

1. 220kV, 330kV, 500kV and 500kV± (see figure 3.30)

2. 330kV, 500kV and 500kV±

3. 500kV and 500kV±

However, these are transmission voltages which would only connect directly to high powered hydroelectric plant (e.g. >50MW). Sub 50MW schemes would require connection to sub transmission voltage network and those less than 10MW would require connection to distribution voltage.

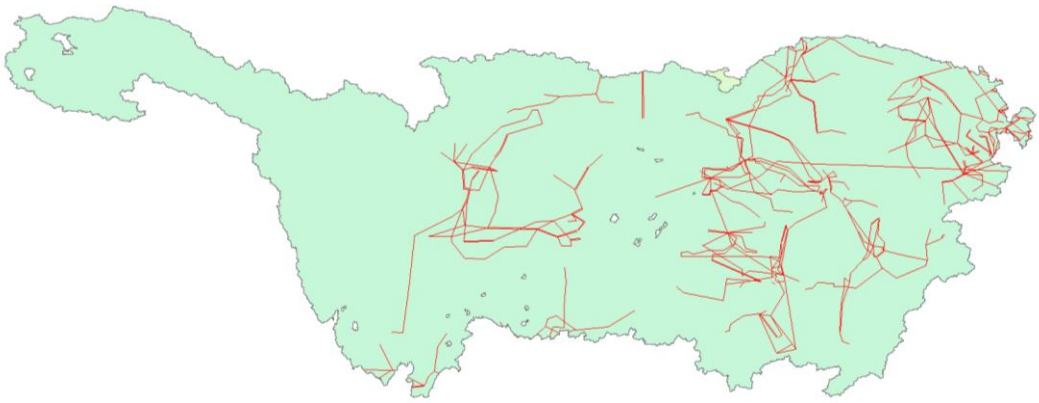


Figure 3.30 DIVA-GIS 220kV, 330kV, 500kV and 500kV± and up transmission network within the Yangtze drainage basin (YDB) (DIVA-GISa, 2016)

As the transmission network dataset is incomplete, a proxy dataset of points was used representing villages, towns and cities (DIVA-GIS, 2016). It is assumed micro-hydropower (<100kW) sites can connect to villages, towns or cities; mini and small hydropower (<50MW) sites can connect to towns and cities only and projects greater than 50MW can connect to cities only. The proxy dataset of villages, towns and cities was filtered by the author to produce a second dataset of towns and cities only and a third of cities only (see figures 3.31; 3.32; 3.33). This is far from ideal and may over- or under-estimate transmission line distances, and hence scheme costs, and an obvious weakness of the analysis.

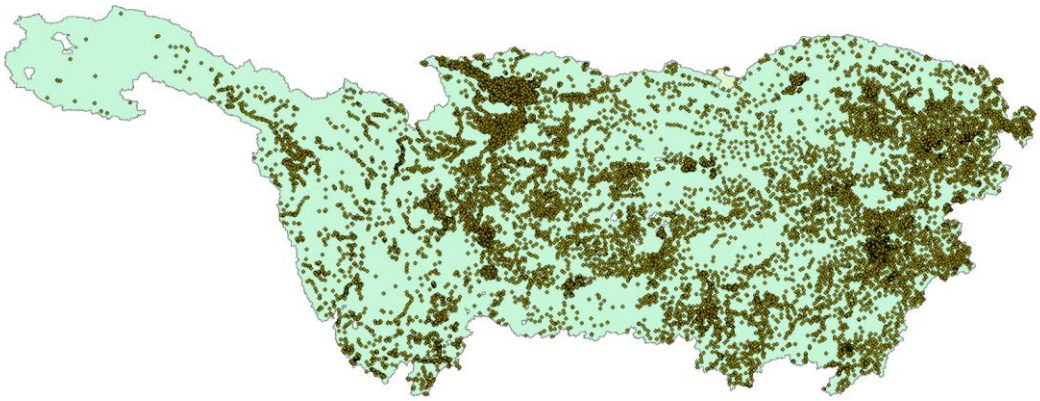


Figure 3.31: DIVA-GIS cities, towns and villages within the Yangtze drainage basin (YDB) dataset (DIVA-GIS$_b$, 2016)

Figure 3.32: DIVA-GIS cities and towns within the Yangtze drainage basin (YDB) dataset (DIVA-GIS, 2016)

Figure 3.33: DIVA-GIS cities only within the Yangtze drainage basin (YDB) dataset (DIVA-GIS, 2015)

3.10 Selection of a power production and costing model

Costing of hydropower is challenging due to the number of contributing suppliers and the fact that hydropower schemes are very site specific. Large damming projects have many uncertainties and there is overwhelming evidence that budgets are systematically biased below actual costs (Ansar et al., 2014).

RETScreen is a clean energy management software suite produced by Natural Resources Canada (RETScreen, 2004) aimed at renewable energy feasibility analysis, including hydropower. The costing methodology is based on an excel based tool called HYDROHELP (Hydrosys Consultants, 2008) developed by Gordon (2008). In its current form the

RETScreen software would only be useful for hydropower costing of single sites within the Yangtze basin, but the equations used within the software were published in a manual entitled Small Hydro Project Analysis (Natural Resources Canada, 2004b). As the software manual details equations for both energy production of a site and costing of each component of a hydropower scheme (for schemes less than 50MW), and the methodology is tried and tested, it was selected as the basis for the power production and costing model used in the hydropower search algorithm. The implementation of this model is discussed in Chapter 6.

3.11 Chapter Conclusions

- The chapter identified the requirement data to search for run-of-river hydropower schemes across a catchment.
- There are several public domain DEM datasets available covering Asia, and HYDROSHEDS was selected due to it being based on a high-quality underlying DEM (SRTM) and designed specifically for hydrological applications.
- The HydroSHEDS DEM at a resolution of 15-arcseconds was selected as a compromise between model quality and model speed.
- The Yangtze drainage basin (YDB) catchment was identified (by the author) with an estimated area of approximately 2 million km^2.
- A river data set across the YDB was established (by the author) with a minimum catchment of $2km^2$, assuming rivers with lower catchment areas would be unsuitable for hydropower development.
- To validate the river dataset, the model derived area of a few catchments were compared to the published area, with error generally less than 3%.
- Further validation was carried out by comparing the location of mainstream confluences to those visually identified in Google Earth, with 85% of confluences within two HydroSHEDS model cells distance. The other 15% were mainly due to early break through of ridges between parallel rivers due to the averaging process of the HydroSHEDS model and hence location of schemes may need verification in some cases.
- Comparison of DEM height of confluences found in Google Earth and the HydroSHEDS DEM resulted in a mean error of 21m higher, particularly pronounced

in the mid-Yangtze regions characterised by steep valleys. This may lead to over/under exaggerated head and may need further corrective action at a later stage.

- A Yangtze lakes and reservoir dataset was created (by the author) by combining multiple freely available GIS datasets within the public domain.
- River flow variation data across all rivers within the Yangtze does not exist and therefore a hydrological model is required to estimate river flow across the river network identified (see chapter 4 and chapter 5).
- Rivers identified are always once cell width, and hence depth and width required for civil works costs estimations can only be estimated using published equations in conjunction with derived river flow data.
- An access road dataset was created (by the author) by combining five freely available GIS datasets within the public domain.
- A major limitation of the datasets required is lack of a transmission network except for voltages above 220kV. A proxy dataset was developed based on locations of villages, towns and cities within the YDB, but it is accepted this may over or under estimate transmission distances. This should be born in mind for future development of this work and cost sensitivity of transmission distances should be carried out (see chapter 7).
- The RETScreen small hydro model was selected as the most thorough costing and power production model available, although it is designed specifically for schemes less than 50MW and therefore it may not be appropriate for larger schemes.

4 Development of a hydrological model: Meteorological data inputs

4.1 Introduction

As introduced in section 3.6, understanding river flow across the YDB is a key component of configuring and costing hydropower sites. Hydropower sites are configured for a specific design flow and appropriate electro-mechanical equipment selected to maximise efficiency and energy production whilst minimising costs. However, river flow data is limited to a few specific river points and absent across most of the YDB. Therefore, it is necessary to develop a hydrological model to derive surface runoff (river flow) conditions across the river network at a sufficient temporal resolution. Hydrological models require terrain data (see chapter 3) and meteorological data which vary both spatially and temporally. Water enters drainage basins through precipitation, falling as rain or snowfall, where snowfall, snow accumulation and snowmelt are largely dependent on temperature. Most of the fallen precipitation flows to the catchment outlet (in this case the Yangtze estuary), although a proportion is returned to the atmosphere via evapotranspiration before reaching the outlet.

Meteorological information is increasingly available as gridded datasets, made available through interpolation of point source meteorological station data, satellite data or derived from meteorological/climate models. Interpolation/modelling and gridding of meteorological point data provides data across whole catchments including ungauged areas. This chapter will focus on development of Yangtze specific gridded precipitation, evapotranspiration and temperature datasets and selection of a snowfall/accumulation and melt model.

4.2 Development of precipitation datasets

4.2.1 Precipitation data collection

Precipitation is condensation falling under gravity due to saturation of the atmosphere with water vapour, and can broadly be defined as rainfall (rain, drizzle), snowfall (snow, graupel, hail) and sleet (precipitation composed of both rain and snow). Precipitation is part of the

hydrological cycle (or water cycle) which describes the movement of water on, above and below the Earth's surface, and responsible for the planet's fresh water. The very nature of precipitation, and observation limitations, make quantification challenging due its fractal nature in space and discontinuity in time (NCAR, 2016).

Historically, precipitation has been measured by the collection of rain/snow in gauges, effectively cylinders which are manually inspected, and the precipitation amount measured. Increasingly, the precipitation can be measured electronically, for example by allowing a measured amount of water to fall through a laser diode and reported automatically to an information network. Gauge data is effectively point data and requires interpolation between networks of gauges to provide coverage over a wide area.

Since the end of the 1970s, satellite data has been increasingly used to create datasets using infrared or microwave measurement. Often satellite data is combined with gauge data to control biases within the satellite data. Sensors from multiple satellites working in combination can lead to a high-quality dataset with high temporal resolution. Regional topography variations can affect precipitation significantly and therefore is heterogeneous in nature even over relatively small catchment areas.

Reanalyses combine historic satellite or gauge observations together with numerical models to simulate aspects of the Earth system and combined objectively to generate a synthesised estimate of the state of the system (Reanalyses.org, 2016), and used extensively in climate research.

4.2.2 An overview of interpolation techniques

Traditionally, linear interpolation techniques were used to map meteorological variables between gauging stations, including drawing of isolines based on the researcher's knowledge (Burrough and McDonnell, 1998). As computing power increased, automated methods were developed including inverse relatively simple distance weighting (IDW), trend surface analysis and Thiessen polygons methods to more complex methodologies including kriging and thin plate splines.

IDW is a deterministic methodology where the assigned values to unknown points are a weighted average of the known points (Shepard, 1968). Trend surface analysis fits a polynomial surface by least-squares regression through the known data points, but is susceptible to error due to outliers in the data (Agterberg, 1964). Thiessen polygons consists

of points in the Euclidean plane whose distance to a known point is less than or equal to every other known point (Thiessen, 1911). Kriging, or Gaussian process regression, is a geostatistical methodology which interpolates values by a Gaussian process governed by prior covariances. Kriging assumes a spatial correlation between the distance or direction of points and fits a mathematical function to several points to determine the output for each location (Matheron, 1960). Thin plate splines are analogous to bending a thin sheet of metal over the known points yet retaining a degree of smoothness (Duchon, 1976).

Evaluation of interpolation methodologies show kriging, particularly co-kriging which includes geographical data, produces the best results but is difficult to implement (Hofstra et al., 2008; Luo, 2007; Haberlandt, 2007).

4.2.3 Selection of a precipitation dataset

In selecting a precipitation dataset, one that is available in the public domain at global or continental scale is preferable, as to enable the dataset to be used to study other catchments. Ideally, the dataset should have a reasonably high temporal resolution, at least daily or sub-daily. Sub-daily measurements are more representative of nature as precipitation rarely falls steadily throughout a day, and averaging of precipitation may unrealistically represent the production of surface runoff within a hydrological model. An intense but short storm can quickly saturate ground producing runoff whereas a long period of steady rain may not.

The National Center for Atmospheric Research (NCAR, 2016) gives a summary of global/continental precipitation datasets on their Climate Data Guide website, split into three broad categories: gauge, satellite and gauge/satellite datasets. Those datasets that cover the Yangtze basin are reproduced in table 4.1 together with reanalysis datasets not on the original list (available from another section of the NCAR Climate Data Guide).

Dataset	Source	Domain	Period of record	Highest temporal resolution	Highest spatial resolution (degrees)	Measurement technique
APHRODITE	University of Tsukuba, Japan Meteorological Agency	Asia	1950 to 2007	Daily	0.25*0.25	Gauge
CFSR	National Centers for Environmental Prediction (NCEP)	Global	1979 to 2011	Sub daily	0.5*0.5	Reanalysis
CHOMPS	Cooperative Institute for Climate and Satellites	Global	1998 to 2007	Daily	0.25*0.25	Satellite
CMAP	National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Centre (CPC)	Global	1979 to 2016	Pentad	2.5*2.5	Satellite/gauge
CMORPH	NOAA CPC	Global	2002 to 2016	Sub-Daily	0.25*0.25	Satellite
COREv2	NCAR	Global	1949 to 2006	Monthly	?	Reanalysis
CPC	NOAA CPC	Global	1979 to 2005	Daily	0.5*0.5	Gauge
CRU	Climate Research Unit	Global	1901 to 2012	Monthly	0.5*0.5	Gauge
ERA-15	European Centre for Medium-Range Weather Forecasts (ECMWF)	Global	1979 to 1993	Sub daily	2.5*2.5	Reanalysis
ERA-20C	ECMWF	Global	1900 to 2011	Sub daily	125km	Reanalysis
ERA-Interim	ECMWF	Global	1979 to 2016	Sub daily	0.75*0.75	Reanalysis
ERA40	ECMWF	Global	1957 to 2002	Sub daily	2.5*2.5	Reanalysis
GHCN-D	NOAA National Climatic Data Center (NCDC)	Global	1880 to 2016	Daily	Station records	Gauge
Global precipitation and temperature	University of Delaware	Global	1900 to 2014	Monthly	0.5*0.5	Gauge
GPCC	World Meteorological Organization	Global	1900 to 2010	Monthly	0.5*0.5	Gauge
GPCP (Daily)	NASA Goddard Space Flight Center (GSFC)	Global	1996 to 2015	Daily	1*1	Satellite gauge
GPCP (Monthly)	NASA Goddard Space Flight Center (GSFC)	Global	1979 to 2015	Monthly	2.5*2.5	Satellite / Gauge
HOAPS	University of Hamburg	Global	1987 to 2008	Sub daily	0.5*0.5	Satellite
JRA-25	Japanese Meteorological Agency	Global	1979 to 2004	Sub daily	2.5*2.5	Reanalysis
JRA-55	Japanese Meteorological Agency	Global	1979 to 2004	Sub daily		Reanalysis
NASA MERRA	NASA	Global	1979 to 2016	Sub daily	0.5*0.667	Reanalysis
NCEP Reanalysis R2	NCEP	Global	1979 to 2016	Sub daily	2.5*2.5	Reanalysis
NEWS	NASA GSFC	Most of globe	1957 to 2016	Sub daily		
NOAA 20C reanalysis	NCEP / NCAR	Global	1850 to 2014	Sub daily	2*2	Reanalysis

Dataset	Source	Domain	Period of record	Highest temporal resolution	Highest spatial resolution (degrees)	Measurement technique
PERSIANN-CDR	UCI Center for Hydrometeorology & Remote Sensing	Global	1983 to 2016	Sub Daily	0.25*0.25	Satellite/gauge
PREC/L	NOAA	Global	1948 to 2016	Monthly	0.5*0.5	Gauge
SSM/I, SSMIS	NASA's Pathfinder Program	Global	1987 to 2015	Daily	0.25*0.25	Satellite
SSMI V7	NASA Remote Sensing Systems	Global		Sub daily		Satellite

Table 4.1: List of Asian and global precipitation datasets using gauge, satellite, satellite/gauge and reanalysis measurements (NCAR, 2016). Datasets with high spatial and temporal resolution highlighted [1]

Of the datasets listed in table 4.1, there are several that are attractive for further consideration due to their high spatial and/or temporal resolution.

1) APHRODITE (Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation) daily gridded precipitation dataset is a continental-scale (Asia) product based on a dense network of daily rain gauge data for Asia, including the Himalayas, available at 0.25° x 0.25° (approximately 23.65km east-west and 27.8km south-north at the Yangtze mouth) with a long time-series dating from 1950 to 2007 (Yatagai et al. 2009; 2012).
2) With the same spatial resolution, PERSIANN-CDR (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Climate Data Record) dates from 1983 to 2016 and is of global reach (Ashouri et al., 2015). Operational PERSIANN is continually kept up to date and available at sub-daily temporal resolution (1-hourly, 3-hourly and 6-hourly) since 2000. Since 2003, PERSIANN-CCS (Cloud Classification System) produces sub-daily resolution. precipitation data at 0.04° x 0.04° resolution (~4km x 4km) and is kept up to date.
3) NASA's MERRA (Modern Era Retrospective-Analysis for Research Applications) (Rienecker et al., 2011) dating from 1979 to 2016 was also selected for further investigation. Although the resolution (0.5° x 0.667° - approximately 47.2km x

[1] (for acronyms see https://climatedataguide.ucar.edu/climate-data/precipitation-data-sets-overview-comparison-table and https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables)

74.1km at the Yangtze mouth) is coarse, this reanalysis product offers data at hourly temporal resolution and other climate data which could be used in evapotranspiration calculations.

4) Similarly, the NCEP CSFR (Climate Forecast System Reanalysis) (NCEP CSFR, 2016) was also considered attractive at 0.5° x 0.5° degrees (approximately 47.2km x 55.6km at the Yangtze mouth), also with sub-daily data.

Although the sub-daily data of the reanalysis products is attractive, it was decided to reject the MERRA and NCEP CSFR options due to their low spatial resolution. Selecting products with daily data is a compromise between model accuracy and model efficiency, due to the lower data processing requirements. Comparison of gauge-based precipitation datasets with reanalyses reveal generally better performance of the gauge-based sets (Katiraie-Boroujerdy et al., 2017; El Kenawy, 2015) and reanalysis performance is often evaluated against the APHRODITE product with the view that the APHRODITE product is the reference standard (Shah and Mishra, 2014; Peña-Arancibia, 2013; Ceglar et al., 2017).

The PERSIANN-CCS product is initially attractive due to its hourly resolution and very high spatial resolution, but generating FDC requires long time-series of data. Furthermore, Yangtze basin river calibration data in the public domain (e.g. from GRDC – see section 3.6) is generally only available prior to 2000 and the PERSIANN-CCS dataset only begins in 2003. In a comparison of high-resolution precipitation datasets to local daily rainfall gauge observations in Urmia Basin, Iran, APHRODITE performed best with only 3% overestimation compared to 26% underestimation by the PERSIANN dataset (Ghajarnia et al., 2015). In an evaluation of six high-resolution satellite and ground-based precipitation datasets over Malaysia, again APHRODITE performed best as PERSIANN overestimated by 2.1% (Tan et al., 2015) - this level of performance is also reflected in other studies (Qi et al., 2016; Vu et al., 2012; Krakauer et al., 2013).

Therefore, due to the high quality of APHRODITE, its relatively high spatial resolution and long time-series, it was selected as the precipitation product for the hydrological model. APHRODITE precipitation data could also be used in studies of other catchments across Asia. PERSIANN-CDR would be a good alternative to investigate non-Asian catchments, and the PERSIANN-CCS dataset particularly interesting in future refinement of the model if post-2003 river point flow data ever becomes available. Note that the APHRODITE precipitation

data was used ‘as provided’, relying on quality control during production of the data-set and reliability assessments available within the wider literature.

4.2.4 Construction of the APHRODITE precipitation dataset

The following actions were carried out by the APHRODITE team. APHRODITE data was interpolated using between 5,000 and 12,000 meteorological stations for any one day, representing 2.3 to 4.5 times the data available through the Global Telecommunications System (GTS) network used to develop most daily precipitation products. One motivation to develop a high-quality rain gauge based precipitation dataset was to help in validation of satellite data and meteorological models, which can be used in both hind and forecasting. Gauge data within the GTS is rather limited over Monsoon Asia, despite its high population and there is a great need for environmental modelling and prediction.

Three categories of data were used in APHRODITE construction. GTS-based data, pre-compiled datasets and individual data collected by the APHRODITE project, sourced from meteorological and hydrology organisations within 26 countries including China. Pre-compiled data used is shown in table 4.2.

Most national and meteorological stations prohibit release of raw data, and hence despite APHRODITE creating gridded products at 0.05° resolution, datasets are only released to the public at 0.25°/0.5° resolution. Following data collection, the data was reformatted and quality controlled. The interpolation algorithm is as follows (Yagatai, 2009):

1. Define monthly precipitation climatology from daily and monthly observations at each station.
2. Compute the ratio of (1) to the WORLDCLIM.
3. Interpolate (2) onto a 0.05-degree grid using the weighted mean method based on Spheremap (Willmott et al. 1985).
4. Define daily precipitation climatology by Fourier interpolation with the first six harmonics.
5. Using the daily climatology (4), daily precipitation analysis was conducted by computing the ratio of the daily observation to the daily climatology (4) for the target day.
6. The output of (5) was interpolated onto a 0.05-degree grid using the same method used in (3).

7. The 0.05-degree data (6) was re-gridded to 0.25- and 0.5-degree products using the area- weighted mean.

Precompiled Dataset	Time Resolution	Countries/Areas
Association of Southeast Asian Nations (ASEAN) compendium climate dataset M	Monthly, daily (limited)	Southeast Asia
Carbon Dioxide Information Analysis Center (CDIAC)	Sub-daily	Former Union of Soviet Socialist Republics (USSR)
Food and Agriculture Organization of the United Nations (FAO)	Monthly, climate normal	Global
Global Energy and Water Cycle Experiment (GEWEX) Asian Monsoon Experiment-Tropics (GAME-T) data center	Daily	Southeast Asia
Global Historical Climatology Network (GHCN)	Monthly, daily	Global
National Center for Atmospheric Research Data Archive (NCAR DS)	Daily	India, Thailand
National Climate Data Center (NCDC)-9813, 9814	Daily	Russia (former USSR), Kazakhstan
The Mekong River Commission (MRC)	Daily	Southeast Asia
Global Telecommunications System (GTS) The global summary of day [NCDC/National Oceanic and Atmospheric, Administration (NOAA)	Daily	Global
World Climatology Data (WorldClim)	Climatology	Global

Table 4.2: Precompiled datasets used in the construction of APHRODITE (Yagatai et al., 2012)

In addition to using WORLDCLIM to correct bias of gauge sparse areas, the weighting function considered both horizontal distance and local topographical features such as elevation and mountain slopes which improves the orographic (i.e. relating to form of clouds and rainfall due to mountains) precipitation pattern. Stations in the same 0.05° grid were combined by averaging of the data. A map of station data is shown in figure 4.1. Note how there are several stations is the YDB area, mostly offline measurements, but the network is not so nearly as dense as in India and Japan. Also, measurements on the Tibetan plateau are very limited potentially leading to greater error across high mountain areas of the YDB.

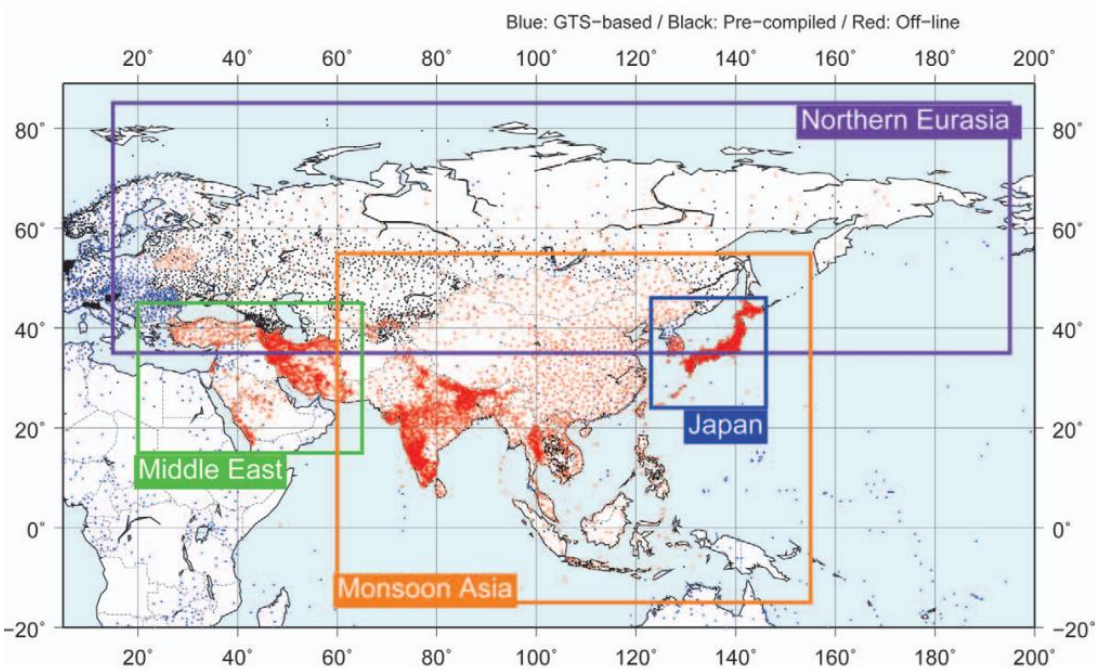


Figure 4.1: Station gauge data used in the construction of the APHRODITE dataset (Yatagai, 2012)

4.2.5 The R programming language

The main software tool employed in this PhD project is the R programming language (R Core Team, 2015) for statistical computing and graphics. All meteorological datasets, the hydrological model and the hydropower search algorithm were developed in R. Microsoft R Open version 3.2.3 (formerly known as Revolutionary R Open) was used to take advantage of multithreaded performance and parallel processing on multi-core PC machines, reducing matrix computing times. R is attractive due to its relatively straightforward programming language and availability of many libraries and packages, which allow statistical techniques and graphical outputs. As R is an interpreted language, there are some speed penalties compared to languages such as C/C+, but the learning curve for such languages is far greater and, for some applications, requires significantly more code to achieve similar results.

4.2.6 Preparation of the APHRODITE data for the hydrological model

APHRODITE V1101 (0.25° resolution monsoon Asia precipitation) data is available from the Data Integration and Analysis System (DIAS), which requires registration. It was decided to use data from 1979 to 2007 as MERRA data used in the creation of the evapotranspiration

dataset (see section 4.3) is only available since 1979. The following actions were carried out by the author of this work.

Data was downloaded as compressed (.gz) NetCDF yearly files (a set of software libraries supporting the creation and sharing of array-oriented data) and imported into the R programming language for processing.

The yearly NetCDF files were read using the ncdf4 library (Pierce, 2017) and each file consists of 365 or 366 (leap-year) time layers on 360 longitude by 280 latitude grids. Within each layer are two variables, precip (daily precipitation analysis interpolated onto 0.25° grids [mm/day]) and rstn (ratio of 0.05° grids with a station [%]). The precipitation data was incorporated (by the author) into a raster stack using the raster library (Hijmans et al., 2016a), effectively 365 (or 366) layers of gridded data with resolution of 0.25°, and the spatial extent from 60° east to 150° east longitude and -15° south to 55° north latitude, projected onto the geographic WGS84 datum. A plot of the 1st layer of year 2000 is shown in figure 4.2 (1st January 2000).

As the gridded terrain datasets created in chapter 3 are at 15-arcseond resolution, the meteorological datasets need to be at the same resolution when running the model. However, it was decided to store the meteorological datasets at the native resolution (0.25°) and increase the resolution within the model as each dataset was required. Although this reduces the efficiency of the model, it minimises data storage. To reduce data storage further, the precipitation rasters were cropped to an area of the YDB. Due to the lower resolution of the precipitation dataset compared to terrain datasets, cropping (or masking) using the YDB polygon could lead to missing cells when the precipitation dataset is increased in resolution to that of the DEM (i.e. 15-arcseconds). Therefore, the precipitation data was cropped to a rectangle around the YDB with an extent of 60° east to 122° east longitude, 22° north to 37° north latitude, and saved as a GEOTIFF raster with the naming convention of 'dd-month-yyyy.tif' (see figure 4.3). GEOTIFF allows geographical information to be embedded within the TIFF image.

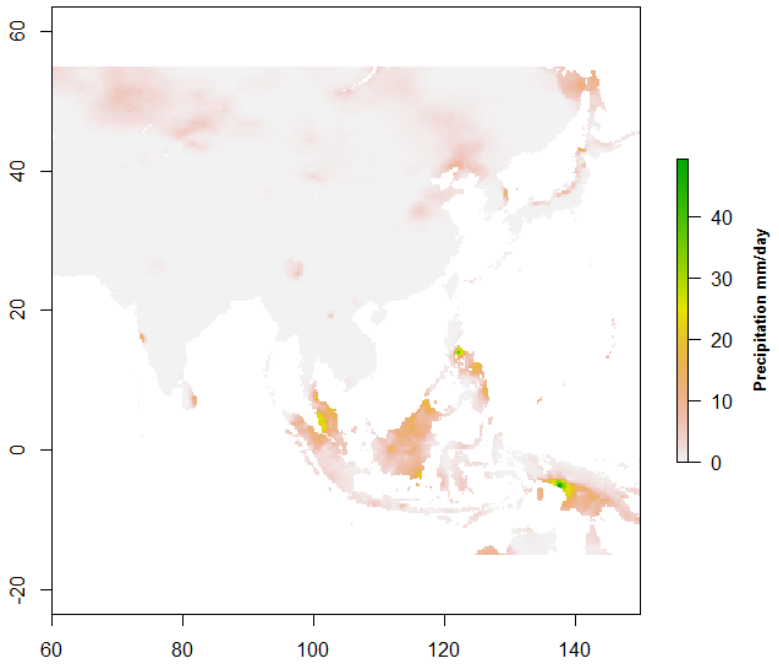


Figure 4.2: Plot of 1st layer of year 2000 (1st January) APHRODITE file after conversion to a raster showing precipitation [mm/day] across Asia (based on data from Yagatai et al., 2012)

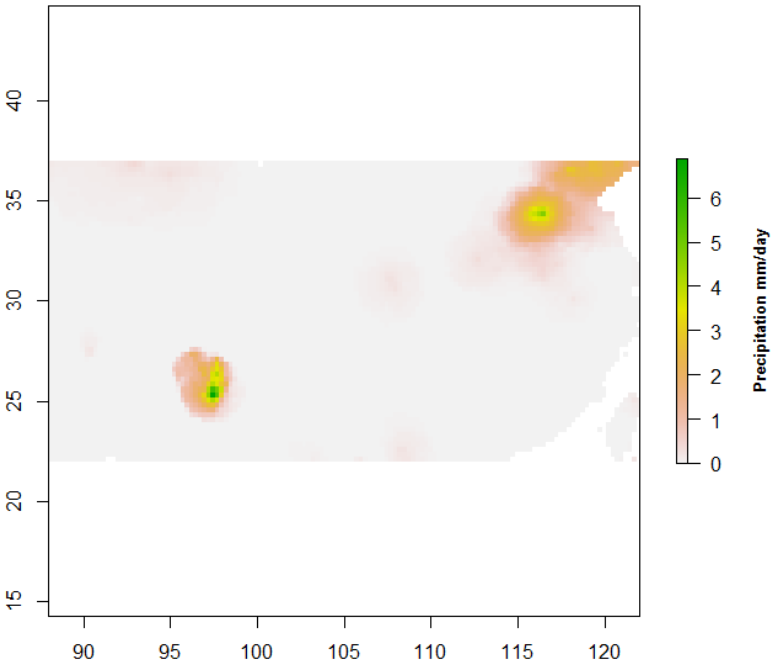


Figure 4.3: Plot of 01-January-2000 APHRODITE precipitation data [mm/day] in a rectangular region around the YBD saved as a GEOTIFF image (based on data from Yagatai et al., 2012)

This process was automated to obtain all days from 1979 to 2007 (29 years; 10,592 days).

4.2.7 Characteristics of the Yangtze basin precipitation within the APHRODITE dataset

To put the precipitation within the YDB in context, figure 4.4 shows a precipitation map across Asia, created by averaging the annual mean precipitation from 1961 to 2007.

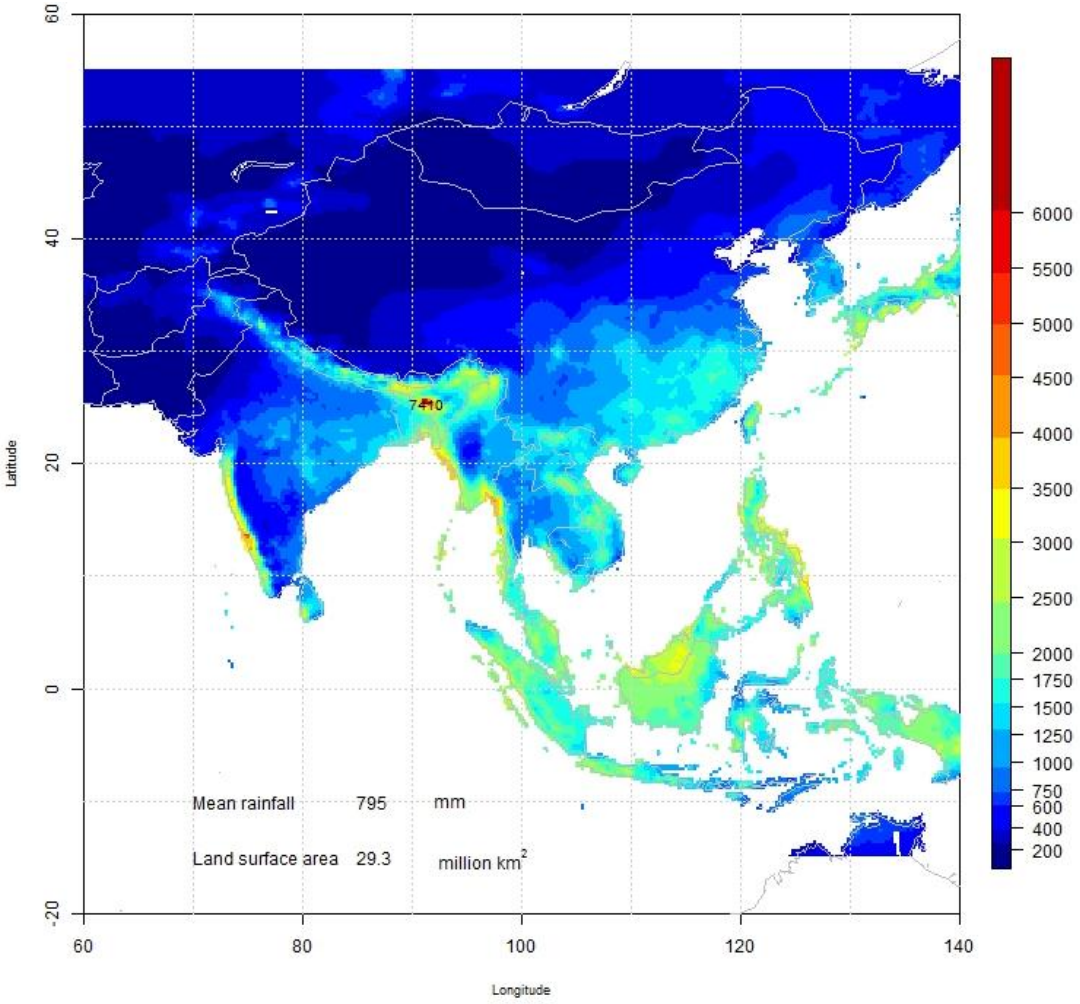


Figure 4.4: Mean annual precipitation [mm/year] across Asia from 1961 to 2007 (based on data from Yagatai et al., 2012)

Overall, China has relatively little precipitation compared to other parts of Asia, particularly compared to the monsoon side of the Himalayas, the Western Ghats of India and the Malay Archipelago, with many areas experiencing less than 1000mm per year. The south-east of China experiences wetter conditions, much of which is within the YDB, with over 2000mm of rain a year in some areas.

To examine characteristics of the YDB further, the datasets were masked using the YDB area polygon. The figures 4.5; 4.6; 4.7; 4.8; 4.9 show annual and seasonal precipitation from 1979 to 2007 across the YDB with mean annual rainfall of 954mm, 251mm in spring (March, April, May), 432mm in summer (June, July, August), 189mm in autumn (September, October, November) and 81mm in winter (December, January, February). Therefore, precipitation is seasonal with wet summers and dry winters, the opposite of UK conditions. Higher precipitation is experienced in the south and east of the basin although maximum seasonal rainfall is experienced in summer in the mid-west. There is very little precipitation in the north-west highlands on the Qinghai/Tibet plateau.

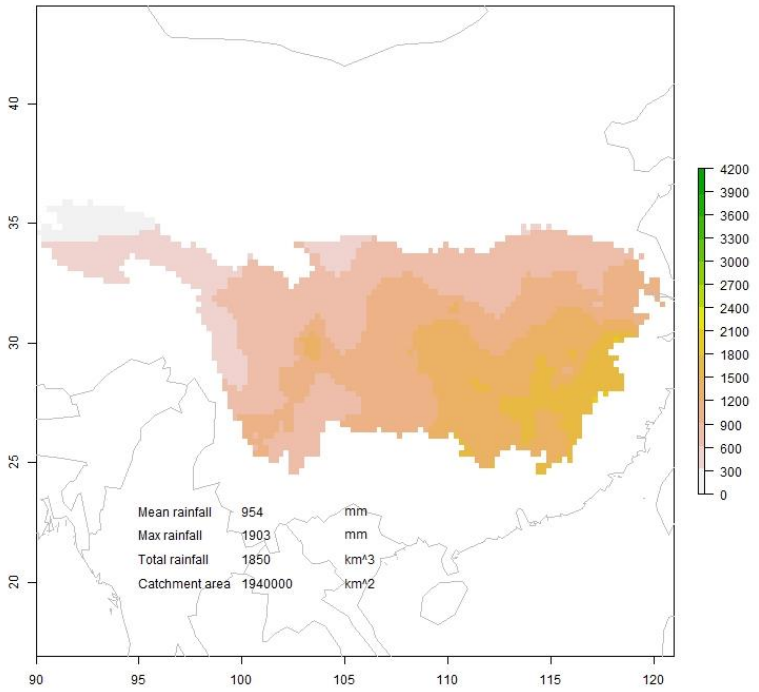


Figure 4.5: Mean annual precipitation [mm yr^{-1}] 1979 to 2007 (based on data from Yagatai et al.,2012)

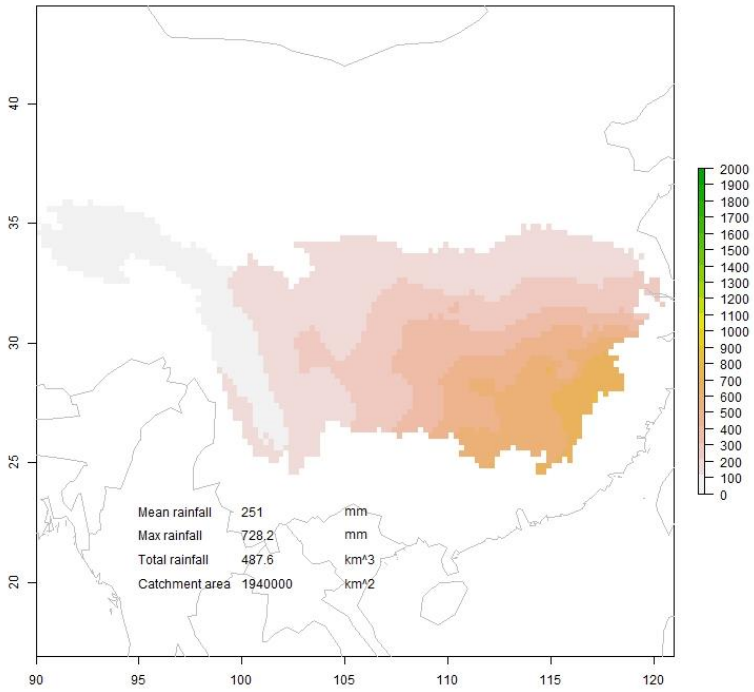


Figure 4.6: Mean spring precipitation [mm season^{-1}] 1979 to 2007 (based on data from Yagatai et al., 2012)

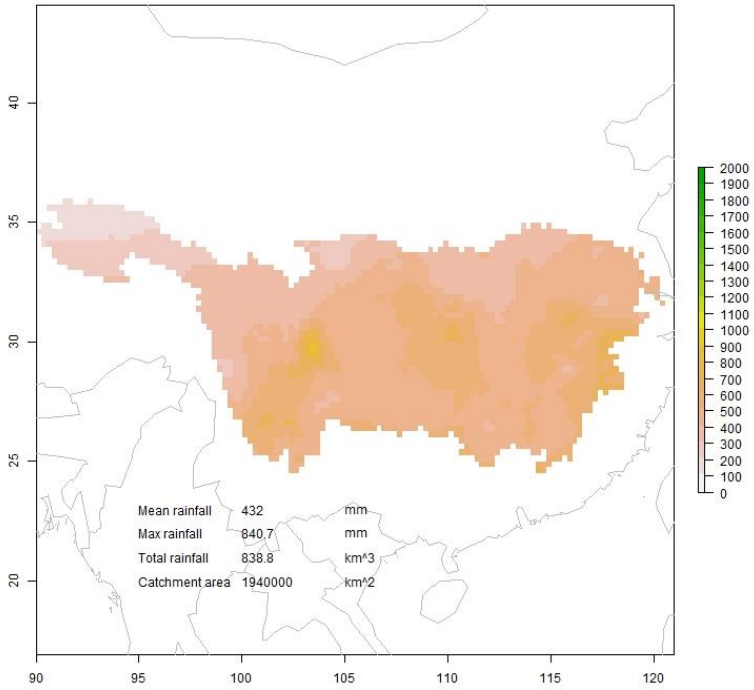


Figure 4.7: Mean summer precipitation [mm season^{-1}] 1979 to 2007 (based on data from Yagatai et al., 2012)

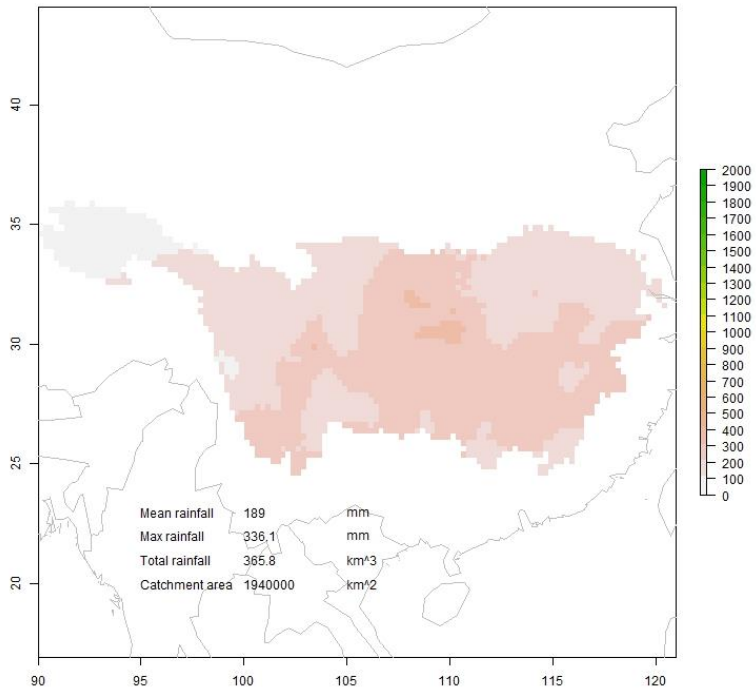


Figure 4.8: Mean autumn precipitation [mm season^{-1}] 1979 to 2007 (based on data from Yagatai et al., 2012)

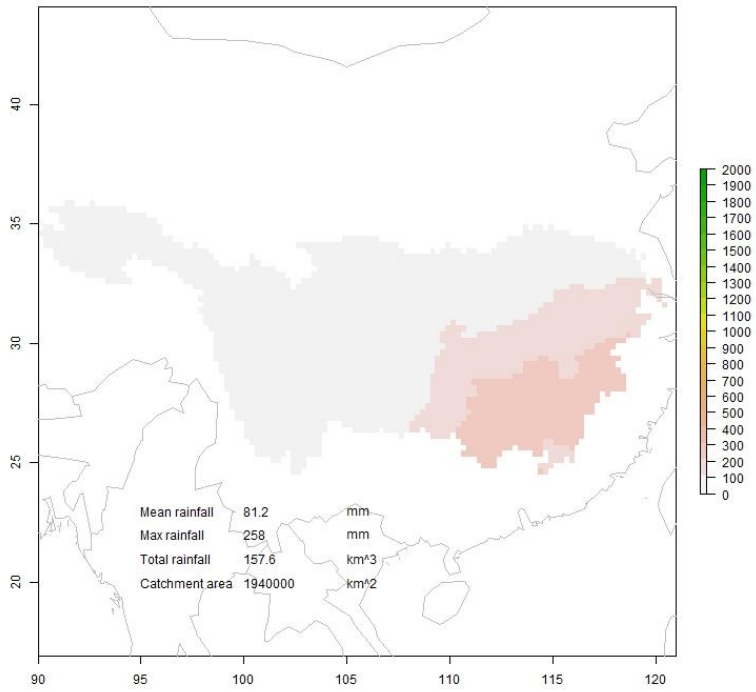


Figure 4.9: Mean winter precipitation [mm season^{-1}] 1979 to 2007 (based on data from Yagatai et al., 2012)

4.2.8 Annual variation in precipitation

As expected across any catchment area, precipitation varies from year to year. Figure 4.10 shows the rolling mean annual precipitation across the basin between 1979 and 2007, with particularly high mean precipitation in 1980, 1983, 1998 and 2002 and low mean precipitation in 1986, 1988, 2001 and 2007. The precipitation across the YDB catchment for the wettest (1980) and driest (1986) years are plotted in figures 4.11 and 4.12 respectively. Overall there is a slight decreasing trend across the precipitation record. Particularly high summer precipitation in 1998 led to disastrous floods, considered the worst Northern China flood in 40 years (see figure 4.13).

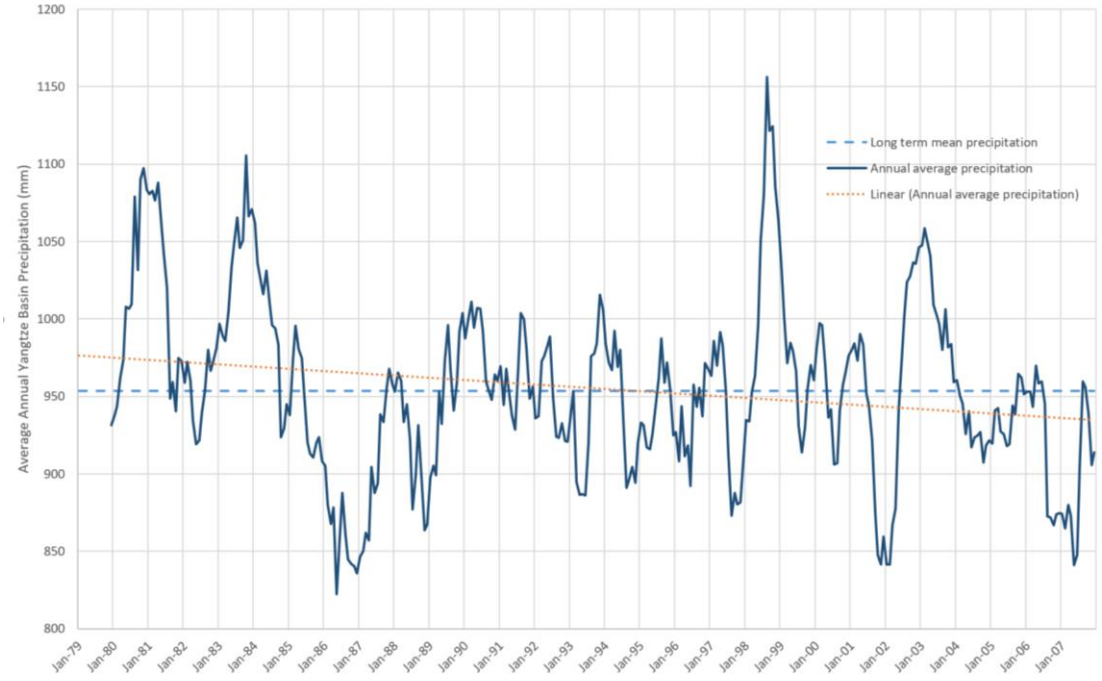


Figure 4.10: Rolling mean precipitation in YDB [mm yr^{-1}] 1979 to 2007 (based on data from Yagatai et al., 2012)

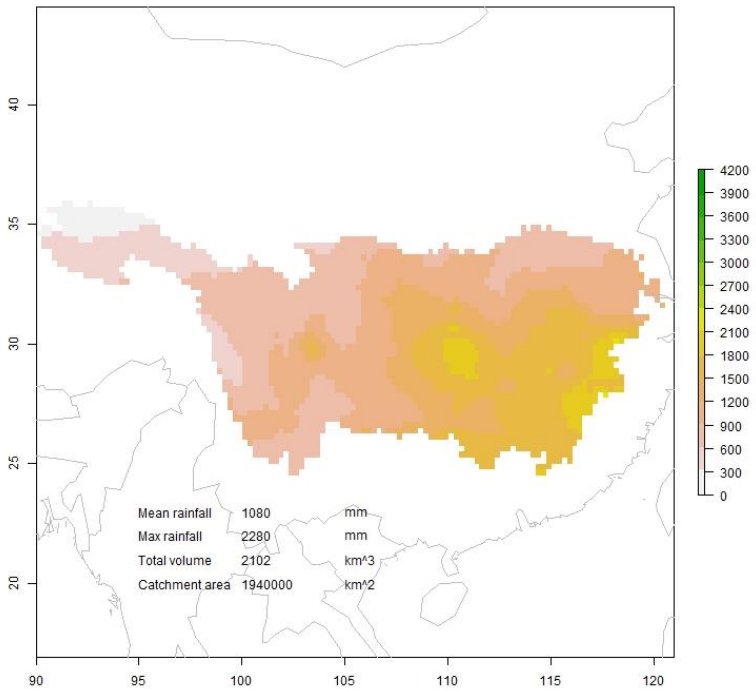


Figure 4.11: Precipitation across the YDB [mm] in 1980, the maximum mean annual precipitation between 1979 and 2007 (based on data from Yagatai et al., 2012)

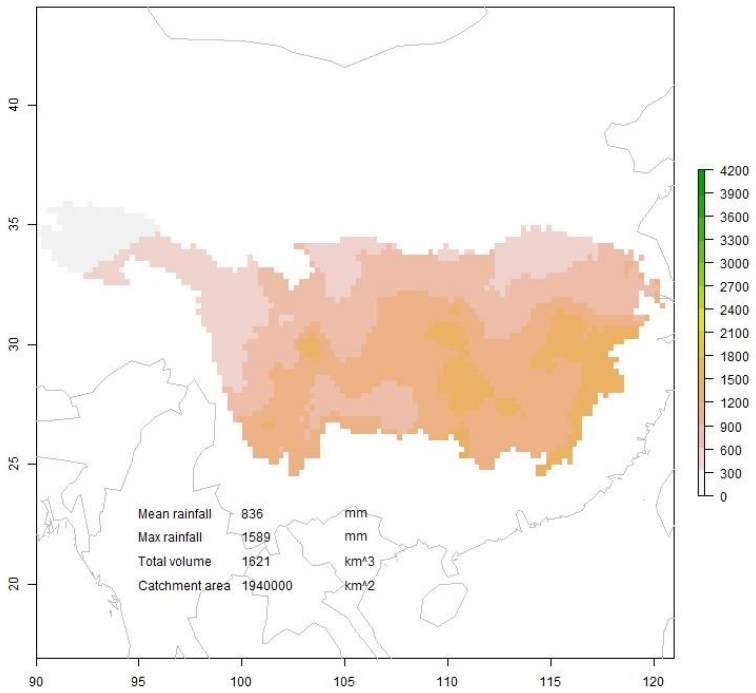


Figure 4.12: Precipitation across YDB [mm] in 1986, the minimum mean annual precipitation between 1979 and 2007 (based on data from Yagatai et al., 2012)

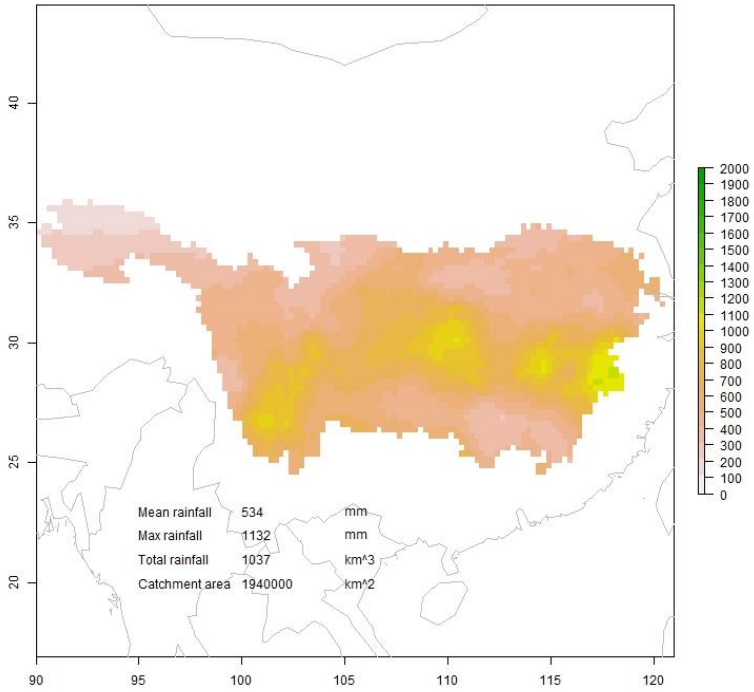


Figure 4.13: Summer (June/July/August) precipitation across YDB [mm] in 1998, resulting in disastrous floods (based on data from Yagatai et al., 2012)

4.3 Development of temperature datasets

4.3.1 Temperature data collection

Temperature is an important variable in runoff modelling as it influences whether precipitation falls as rain or snow (or a mixture of both) and the accumulation and melt of snow. Although the sum precipitation is the same, cold temperatures can delay when runoff occurs due to precipitation lying as snow, particularly at higher altitudes. Temperature is also an important contributing variable to evapotranspiration calculations, the only loss of water from a catchment apart from water flowing out at the catchment outlet. Unlike precipitation, temperature is less fractious in space and changes slowly as latitude changes, assuming elevation, landcover type and weather conditions remain similar. However, temporally, temperature varies diurnally and seasonally, due to changes in elevation and changing weather, and location factors such as proximity to the sea and land-cover.

Temperature is measured via gauge data or via satellite or a mixture of the two. Satellites do not measure temperature directly but infer temperature due to radiances in wavelength bands

(e.g. infrared or microwave). However, the most cited global datasets use a combination of land-station near surface air temperature measurements in conjunction with sea surface temperatures (SST) obtained from ship and buoy measurements (NCAR, 2016). Methodologies like the interpolation of precipitation data are used to create gridded temperature datasets. Atmospheric reanalyses also offer global temperature datasets.

4.3.2 Selection of a temperature dataset

NCAR also offer a list of global temperature datasets, and this is shown in table 4.3 with the addition of the APHRODITE temperature dataset and satellite derived datasets. The reanalyses listed in table 4.1 are also potential candidates.

Dataset	Source	Domain	Period of record	Highest temporal resolution	Highest spatial resolution (degrees)	Measurement technique
AIRS	NASA-JPL	Global	2002 to 2011	Monthly	1*1	Satellite
APHRODITE	University of Tsukuba, Japan Meteorological Agency	Asia	1961 to 2007	Daily	0.25*0.25	Gauge
BEST	Berkeley Earth	Global	1701 to 2013	Daily	1*1 (0.25*0.25 for Europe and USA)	Gauge
Global (land) precipitation and temperature	University of Delaware	Global	1900 to 2014	Monthly	0.5*0.5	Gauge
GISTEMP	NASA Goddard Institute for Space Studies (GISS)	Global	1880 to 2016	Monthly	2*2	Gauge
HadCRUT4 and CRUTEM4	University of East Anglia Climate Research Unit	Global	1850 to 2016	Climatology, Monthly	5*5	Gauge
MLOST: NOAA	NOAA NCDC	Global	1871 to 2016	Monthly	5*5	Gauge
MSU and AMSU	Remote Sensing Systems	Global	1979 to 2012	Monthly	2.5*2.5	Satellite
NCEP-CSFR	NCEP	Global	1979 to 2010	Hourly	0.5 *0.5	Re-analysis

Table 4.3: List of Asian and global temperature datasets using gauge and satellite measurements (NCAR, 2016)[2]. Datasets with high spatial and temporal resolution highlighted

[2] for acronyms see https://climatedataguide.ucar.edu/climate-data/global-temperature-data-sets-overview-comparison-table

A temperature dataset of high spatial and temporal resolution is desirable. As a high-resolution dataset with sub-daily timestep is not available, the APHRODITE daily temperature dataset at 0.25° resolution was once again selected (Yasutomi et al., 2011) for the underlying data for snowfall/accumulation/melt modelling due to the high spatial resolution.

However, for development of evapotranspiration datasets (see section 4.4), minimum and maximum daily temperatures are required along with other meteorological data. The MERRA reanalysis suite has all the required data at high temporal (one-hourly) but low spatial resolution (0.5° x 0.667° degrees). Hence, for evapotranspiration calculations, MERRA temperature data was preferred. NCEP CSFR data was a viable alternative due to its higher spatial resolution (0.5° x 0.5°) and one-hourly resolution, but MERRA was preferred since the data is kept current whereas the NCEP CSFR data is only available up to 2011.

4.3.3 Construction of the APHRODITE and MERRA temperature datasets

This section refers to development of the APHRODITE product by the APHRODITE team. The APHRODITE temperature dataset was based on station data collected for the creation of the precipitation dataset, and the number of stations used estimated to be 1.5-3 times more than the GTS. APHRODITE claims it is the only Asian gridded temperature dataset with both high temporal and spatial resolution. Many of the country meteorological and hydrology organisations contributing to construction of the precipitation dataset also contributed to creation of the temperature dataset. Following quality control, the interpolation algorithm was based on the precipitation interpolation algorithm but did not include monthly mean temperature and climatology, and is described as follows:

1) The temperature was corrected to mean sea level with a temperature lapse rate 6.0°C/km.
2) The corrected temperature was interpolated onto 0.05° grids using a distance-weighting function based on Spheremap, with 100 km of effective radius. The number of observations was also calculated.
3) The temperature was corrected to GTOPO30 elevation with a temperature lapse rate of 6.0°Ckm^{-1}.
4) The 0.05° data was re-gridded to a 0.50° (and subsequently 0.25°) grid product using area-weighted means.

MERRA datasets were generated by applying the Goddard Earth Observing System Data Assimilation System (GEOS DAS) (Suarez et al., 2008) integrating an Atmosphere General Circulation Model (AGCM) with Grid-point Statistical Interpolation (GSI) Analysis. GEOS-5 is a weather-and-climate capable model used for atmospheric analyses, weather forecasts, coupled and uncoupled climate simulations and predictions and coupled chemistry-climate simulations.

4.3.4 Preparation of the APHRODITE and MERRA temperature datasets

This section refers to actions carried out by the author. In a similar manner to the APHRODITE precipitation data, APHRODITE temperature data (APHROTEMP, V1204R1) is also available from the DIAS website from 1951 to 2007 at 0.25° x 0.25° resolution. Data was downloaded from 1979 in NetCDF format and processed in R with the same methodology of the precipitation dataset, cropped to an area around the YDB and stored at the native resolution. Plots (Asia and YDB area) of the 1st layer of the year 2000 nc file (1st January 2000) converted to raster format are shown in figures 4.14 and 4.15.

MERRA data is available for download from the Goddard Earth Sciences Data and Information Service Center (NASA, 2017). Surface temperature data is available within the IAU (Incremental Analysis Update) 2D surface and TOA (top of the atmosphere) radiation fluxes (tavg1_2d_rad_nx) suite, given the label TS (Surface temperature). At the time of developing the evapotranspiration data, the extent of the hydrological modelling was unknown and hence data for the whole of Asia was collected (60° to 140° west to east, -20° to 60° south-north) from 1979 (start of the MERRA data) to 2007 (to match the end of the APHRODITE data). To establish the minimum and maximum temperature of each day, 24 hourly files for each day were downloaded and combined into a raster stack (or brick), and the minimum and maximum value of each cell column extracted. These values were then combined into minimum and maximum temperature datasets, an example given in figure 4.16 and 4.17 for 1st January 2000. Note how maximum diurnal difference across Asia on that day is 25.6°C.

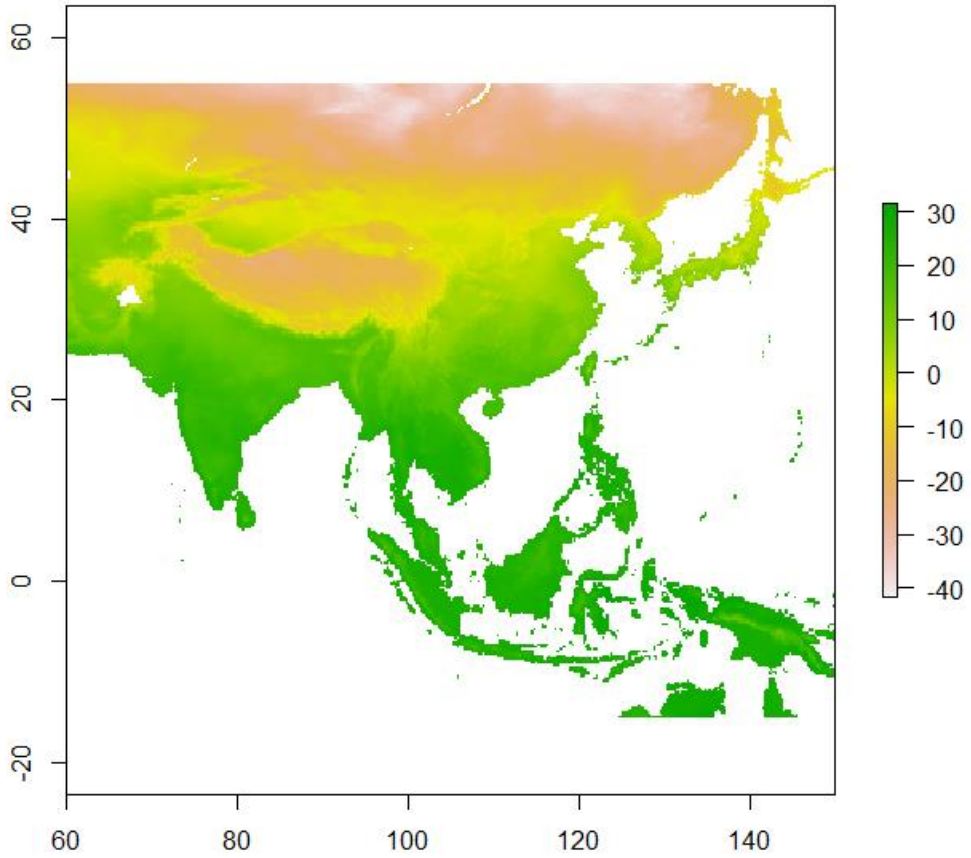


Figure 4.14: Plot of 1st layer of year 2000 (1st January) APHRODITE temperature raster [°C] across Asia (based on data from Yasutomi et al., 2011)

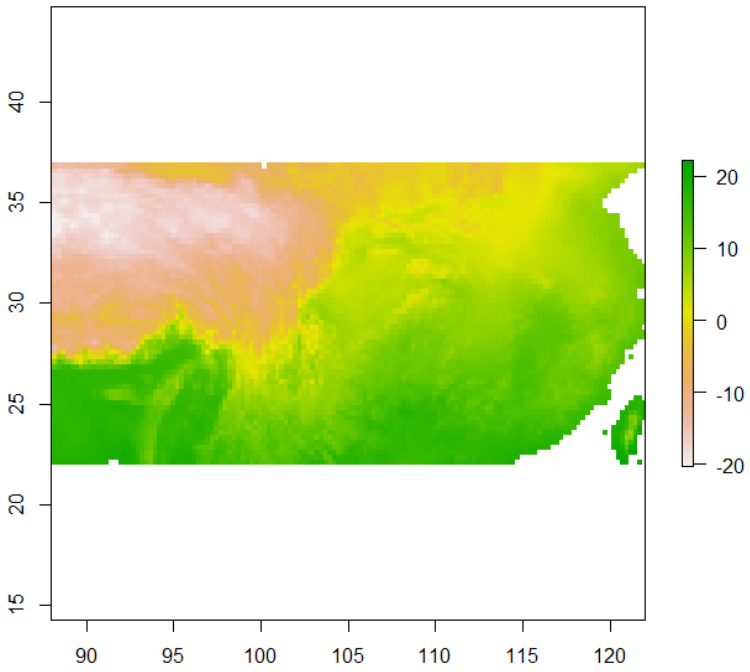


Figure 4.15: Plot of 01-January-2000 APHRODITE temperature raster data [°C] in rectangular region around the YBD (based on data from Yasutomi et al., 2011)

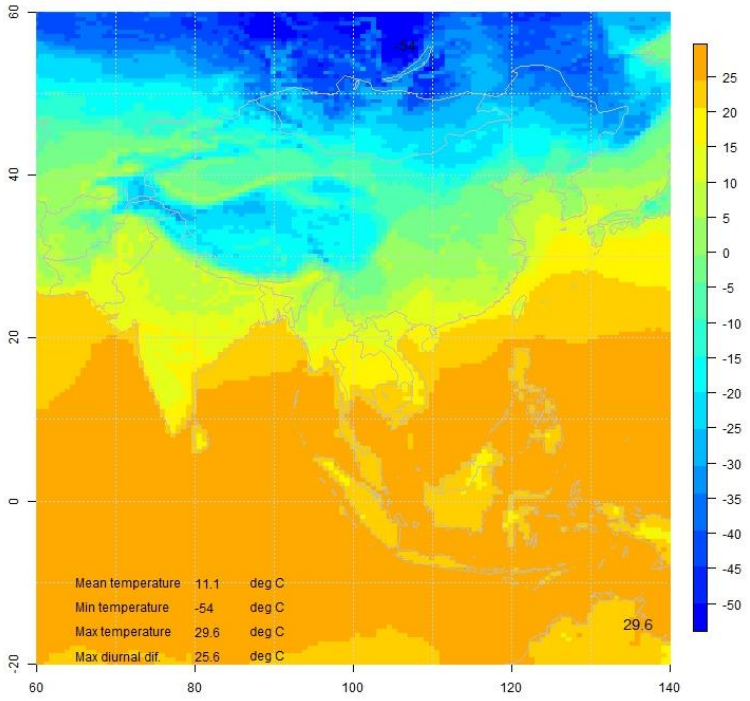


Figure 4.16: 1st January 2000 MERRA minimum temperature raster [°C] across Asia (based on data from Rienecker et al., 2011)

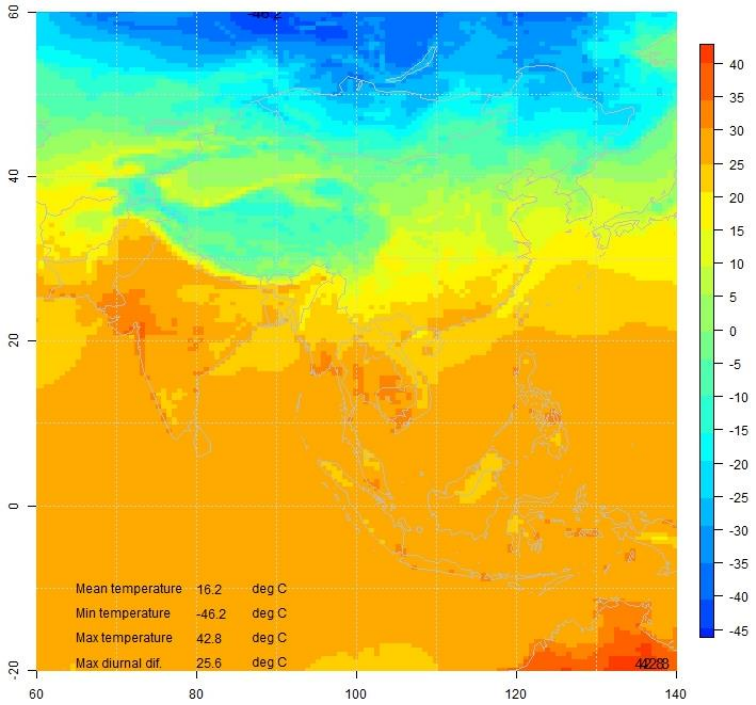


Figure 4.17: 1st January 2000 MERRA maximum temperature raster [°C] across Asia (based on data from Rienecker et al., 2011)

4.3.5 Characteristics of the Yangtze basin temperature within the APHRODITE and MERRA datasets

Across the YDB, the mean annual temperature is 11.2°C, with a maximum of 20.3°C (mean) in the central Sichuan basin and -9.4°C (mean) on the high Qinghai-Tibet plateau. The YDB is characterised by warm mean annual temperatures (>15°C) across the Sichuan basin and eastern third of the catchment, cool temperatures (5 to 15°C) in the mid-Yangtze highlands and cold temperatures in the upper Yangtze (-10° to 5°C) (see figure 4.18). Seasonally (see figures 4.19, 4.20, 4.21, 4.22) the YDB is characterised by hot summers, particularly in the east and Sichuan basin, and cold winters, particularly on the Qinghai/Tibet plateau.

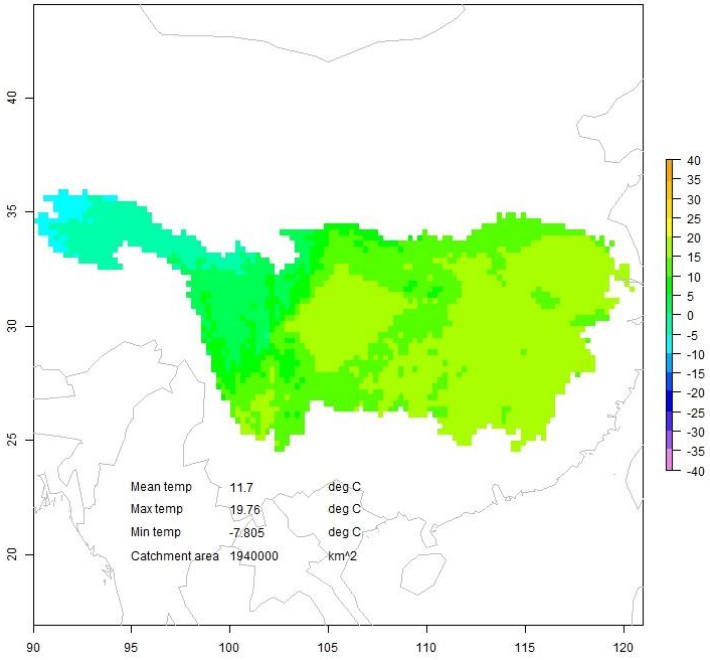


Figure 4.18: APHRODITE mean temperature [°C] across YDB from 1979 to 2007 (based on data from Yasutomi et al., 2011)

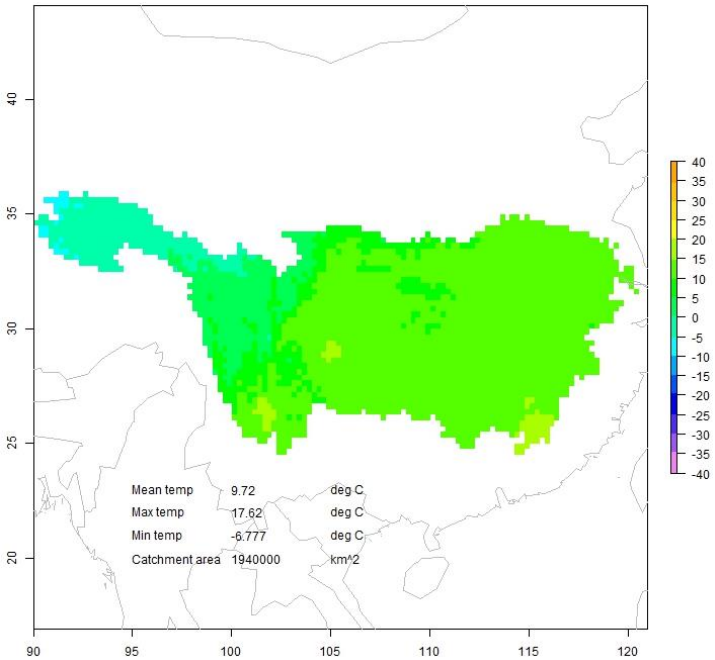


Figure 4.19: APRHODITE mean spring (March, April, May) temperature [°C] across YDB from 1979 to 2007 (based on data from Yasutomi et al., 2011)

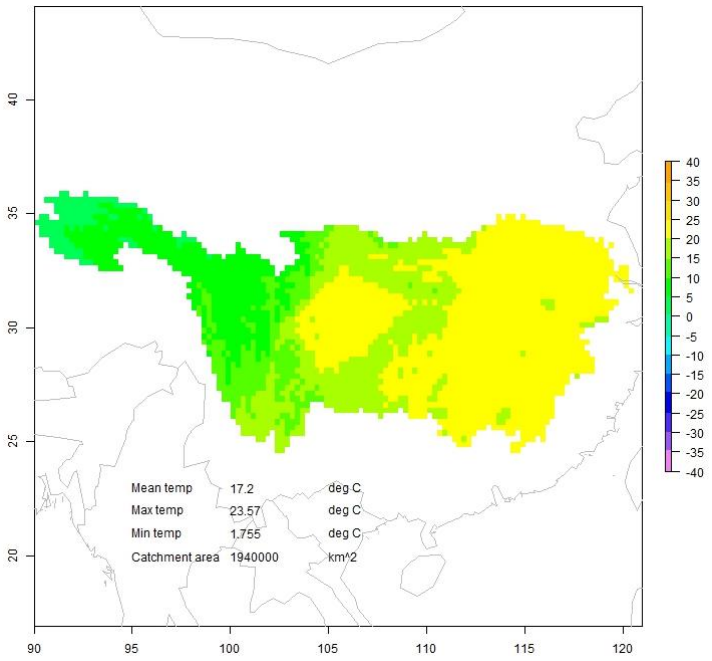


Figure 4.20: APRHODITE mean summer (June, July, August) temperature [°C] across YDB from 1979 to 2007 (based on data from Yasutomi et al., 2011)

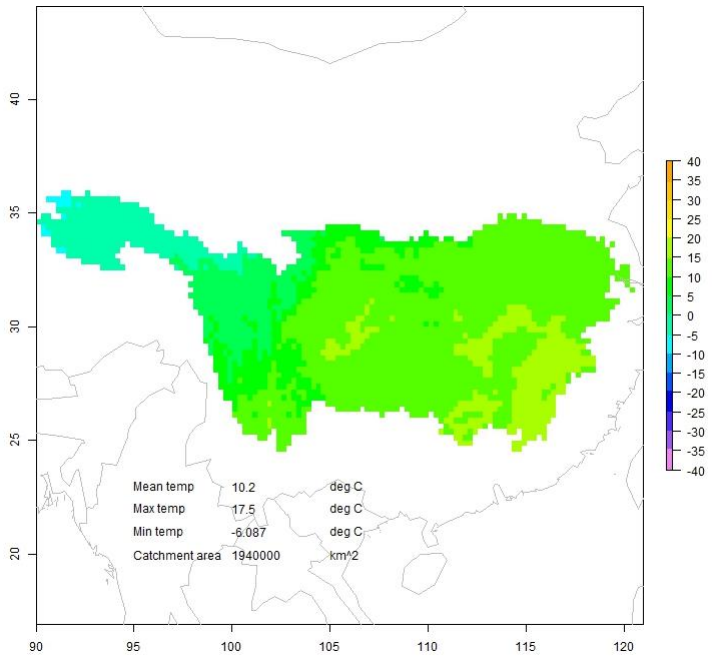


Figure 4.21: APRHODITE mean autumn (September, October, November) temperature [°C] across YDB from 1979 to 2007 (based on data from Yasutomi et al., 2011)

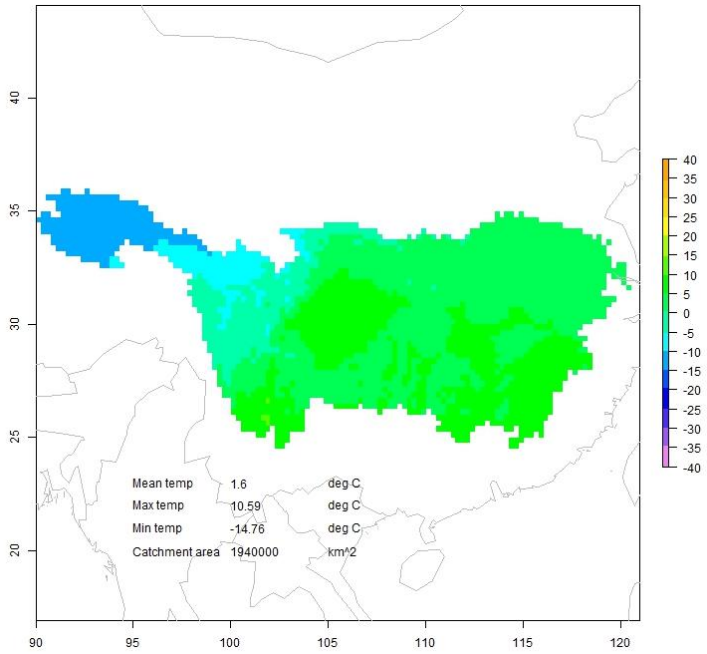


Figure 4.22: APRHODITE mean winter (December, January, February) temperature [°C] across YDB from 1979 to 2007 (based on data from Yasutomi et al., 2011)

The MERRA dataset mean temperature (11.7°C) across the YDB from 1979 to 2007 is approximately 0.5 degrees lower than the APHRODITE dataset (11.2°C) (see figure 4.24).

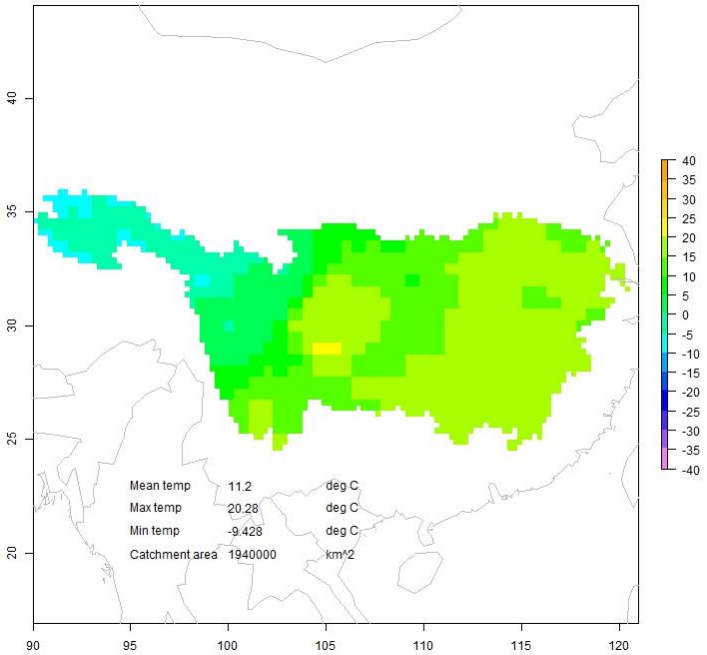


Figure 4.23: MERRA mean temperature [°C] across YDB from 1979 to 2007 (based on data from Rienecker et al., 2011)

4.3.6 Annual variation in the temperature dataset

Temperature also varies annually, potentially increasing/decreasing evapotranspiration rates and changing snowfall and subsequent accumulation and melt. Figure 4.24 shows the rolling mean annual temperature for both the APHRODITE and MERRA datasets. Both datasets show a similar pattern, with the MERRA dataset biased 0.5°C lower. The reasons for this bias are unclear, but Wang (2014) warns that temperature datasets may use the integral of the continuous temperature measurements over a day, or the mean of the daily maximum and minimum temperature measurements, and hence caution should be exercised when using mean temperature datasets. Both datasets show a warming trend of approximately 1°C over the length of the record, possibly indicating climate change which could signify changing future hydropower conditions.

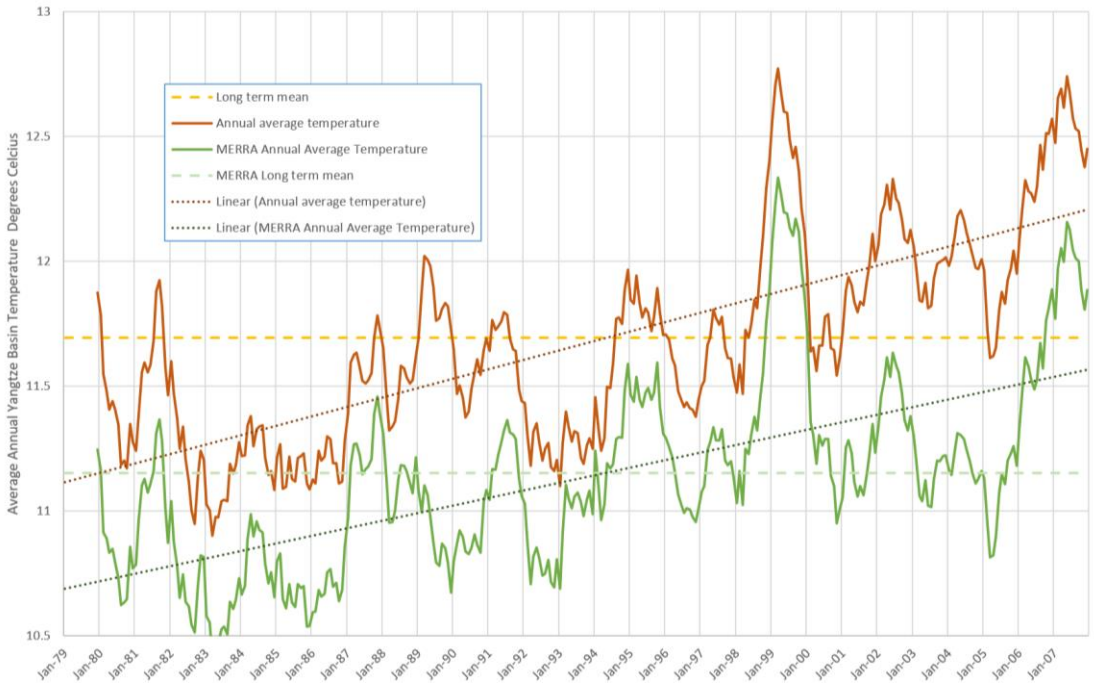


Figure 4.24: APHRODITE (orange) and MERRA (green) rolling mean temperature [°C] across YDB from 1979 to 2007 (based on data from Yasutomi et al., 2011 / Rienecker et al., 2011)

4.3.7 MERRA diurnal variation in temperature

As stated earlier, evapotranspiration calculations rely on minimum and maximum daily temperatures which are only available by employing a sub-daily dataset (e.g. MERRA). Using APHRODITE daily temperatures for snowfall/accumulation and melt is a compromise

between spatial and temporal resolution. Figure 4.25 shows an example of hourly MERRA temperature variation over one day (1st April 2000), with maximum mean temperature of 22.9°C at 1pm and a minimum mean temperature of 6.3°C at 5am/6am, a diurnal range of 16.6°C.

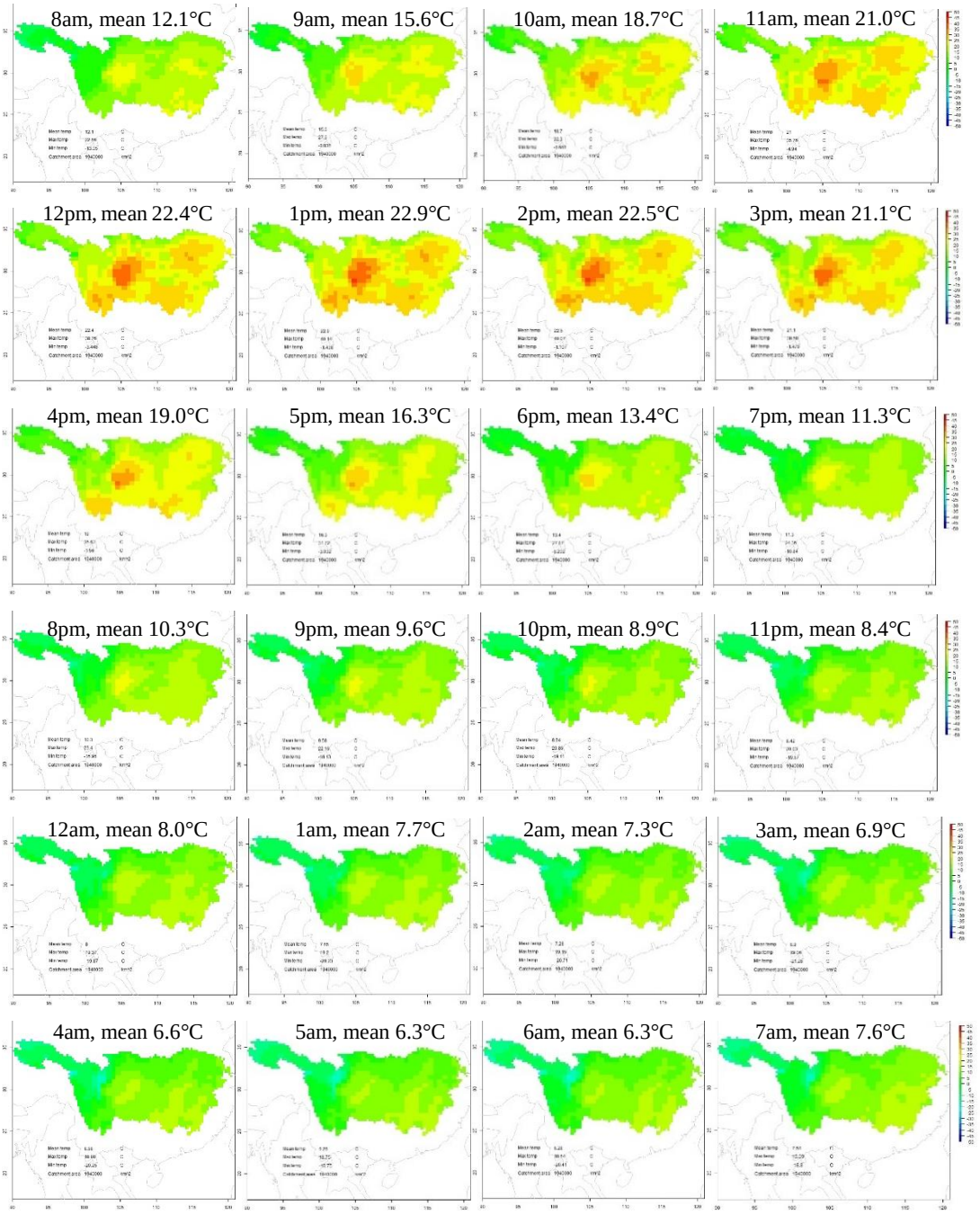


Figure 4.25: MERRA hourly temperature variation (°C) for 1st-April-2000 across the YDB (based on data from Rienecker et al., 2011)

This highlights the compromise made in using higher spatial APHRODITE daily data. For example, in some locations, the average daily temperature may be above the snowfall temperature and the model would assume that precipitation falls as rain, but during the colder hours, precipitation could fall as snow due to temperatures below the snowfall temperature. Therefore, snowfall events may be missed, however using hourly data inputs would require significant data storage.

4.4 Development of evapotranspiration datasets

4.4.1 An introduction to evapotranspiration

Evapotranspiration is the combination of water loss to the atmosphere due to evaporation at the surface and transpiration from crops and vegetation. Evaporation is when water is converted to water vapour (vaporisation) and removed from surfaces including lakes, rivers, paved areas, soils and wet vegetation to the atmosphere. Transpiration is the vaporisation of liquid water in plant tissues and lost to the atmosphere predominately through stomata (small openings on a plant leaf). Both processes occur simultaneously and hence it is difficult to distinguish between the two processes. Evapotranspiration rates change over the growing season, initially dominated by evaporation when vegetation growth is at a minimum, but transpiration becomes the main process as vegetation develops.

The Food and Agriculture Organization of the United Nations (FAO) has developed a standardised methodology to calculate daily evapotranspiration rates, usually employed to calculate crop water requirements (Allen et al., 1998), based on the Penman-Monteith equation (Penman 1948; Monteith 1965). In this project, evapotranspiration loss is an important component in determining net surface discharge. Evapotranspiration rates are mainly determined by the availability of water and the amount of solar radiation reaching the surface. Other weather parameters affecting evapotranspiration rates include air temperature, humidity and wind speed, and the evaporating power of the atmosphere can be calculated by assuming a standardised vegetated surface, or reference crop.

4.4.2 Reference crop evapotranspiration

The evaporating power of the atmosphere can be estimated by assuming a reference crop not short of water (the FAO method uses a hypothetical grass 0.12m tall), called the reference crop evapotranspiration, denoted as ET_o. The only factors affecting ET_o are weather and

climatic parameters, and crop and soil factors are ignored. ET_o is often incorrectly referred to as potential evapotranspiration (PET), but PET depends on crop factors and is calculated as:

$$\boldsymbol{PET = ET_o \times k_c} \quad (4.1)$$

where PET is potential evapotranspiration [mm day^{-1}], ET_o is reference crop evapotranspiration [mm day^{-1}] and k_c is crop coefficient [dimensionless – changes over the growing season].

4.4.3 Available evapotranspiration datasets

Mueller et al. (2011) evaluated global observations-based evapotranspiration datasets, 18 models including diagnostic datasets, land surface models, reanalyses and global climate models. Few of these are available as gridded data dating back to 1979 at daily temporal resolution, which is required to be in keeping with the precipitation and temperature data developed earlier. Global (excluding Antarctica) gridded monthly data is available at 0.5° x 0.5° resolution dating back to 1901 (University of East Anglia Climatic Research Unit, 2013). The Numerical Terradynamics Simulation Group (NTSG) released a global 8-day average dataset using remote sensing data from the MODIS satellite (Mu et al., 2013) with data available only from 2000. NTSG also host a global long-term (1983 to 2006) daily evapotranspiration record (Zhang et al., 2010).

The MERRA Reanalyses Suite has all the climatic and weather variables required to compute ET_o for the YDB at daily resolution dating back to 1979. Section 4.4.4 documents the development of a long-term daily YDB evapotranspiration dataset (by the author) based on data from the MERRA Reanalyses Suite.

4.4.4 Development of a long-term daily YDB evapotranspiration dataset from MERRA data using the FAO methodology

Following the FAO methodology (Allen et al.,1998), evapotranspiration is calculated by an energy balance where the energy arriving at the surface must equal the energy leaving the surface for the same period:

The equation for an evaporating surface can be written as:

$$\boldsymbol{R_n - G - \lambda ET - H = 0} \quad (4.2)$$

where R_n is net radiation [Wm^{-2}], G is soil heat flux [Wm^{-2}], λ ET is the latent heat flux [Wm^{-2}] and H is the sensible heat flux [Wm^{-2}]. Sensible heat is heat exchanged by a body or thermodynamic system in which the exchange of heat changes the temperature of the body or system but leaves unchanged certain other variables of the body or system, such as volume or pressure.

λET represents the evapotranspiration fraction and is calculated from the other variables. R_n and G can be measured or estimated from climatic variables, but H is complex to obtain requiring accurate measurement of temperature gradients. Evapotranspiration can also be determined by various components of the soil water balance from:

$$\boldsymbol{ET = I + P - RO - DP + CR \pm \Delta SF \pm \Delta SW} \qquad (4.3)$$

where I is irrigation, P is rainfall, RO is surface run off, DP is deep percolation, CR is capillary rise, SF is subsurface flow and SW is soil water content.

The Penman-Monteith form of the combination of the soil water balance and energy balance equation is:

$$\boldsymbol{\lambda ET} = \frac{\Delta(\boldsymbol{R_n - G}) + \boldsymbol{p_a c_p}\frac{(\boldsymbol{e_s - e_a})}{\boldsymbol{r_a}}}{\Delta + \boldsymbol{\gamma}\left(\mathbf{1} + \frac{\boldsymbol{r_s}}{\boldsymbol{r_a}}\right)} \qquad (4.4)$$

where λET is evapotranspiration [$MJm^{-2}day^{-1}$], R_n is net radiation [$MJm^{-2}day^{-1}$], G is soil heat flux [$MJm^{-2}day^{-1}$], $e_s - e_a$ is the vapour pressure deficit [kPa], ρ_a is the mean air density at constant pressure [kgm^{-3}], c_p is specific heat of the air [$MJkg^{-1}{}^{o}C^{-1}$] which equals 1.013 x 10^{-3} MJ kg^{-1} C^{-1}, Δ is the slope of saturation vapour pressure temperature relationship [$kPa^{o}C^{-1}$], γ is the psychometric constant [$kPa^{o}C^{-1}$], r_s is bulk surface resistance [sm^{-1}] and r_a is bulk aerodynamic resistance [sm^{-1}].

4.4.4.1 Reference crop calculations for bulk aerodynamic and bulk surface resistances (r_a and r_s)

The bulk aerodynamic resistance (r_a) [sm^{-1}] determines the transfer of heat and water vapour from the evaporating surface into the air above the canopy and calculated as:

$$\boldsymbol{r_a} = \frac{\boldsymbol{ln}\left[\frac{\boldsymbol{z_m - d}}{\boldsymbol{z_{om}}}\right]\boldsymbol{ln}\left[\frac{\boldsymbol{z_h - d}}{\boldsymbol{z_{oh}}}\right]}{\boldsymbol{k^2 u_z}} \qquad (4.5)$$

where z_m is the height of wind resistance [m], z_h is the height of humidity measurements [m], d is the zero plane displacement height [m], z_{om} is roughness length governing momentum transfer [m], z_{oh} is roughness length governing transfer of heat and vapour [m], k is von Karman's constant (= 0.41) [dimensionless] and u_z is the wind speed at height z [ms^{-1}].

The equation is restricted for neutral stability conditions, i.e. where temperature, atmospheric pressure, and wind velocity distributions follow nearly adiabatic conditions (i.e. no heat exchange). For a wide range of crops, the zero-plane displacement height, d, the roughness length governing momentum transfer, $z_{om.}$, and the roughness length governing transfer of heat and vapour, z_{oh}, can be estimated from the crop height h [m] by the following equations:

$$\boldsymbol{d = \frac{2}{3}h} \tag{4.6}$$

$$\boldsymbol{z_{om} = 0.123h} \tag{4.7}$$

$$\boldsymbol{z_{oh} = 0.1 z_{om}} \tag{4.8}$$

For a reference crop of 0.12m (hypothetical grass surface) and standardised height measurements of 2m this yields:

$$\boldsymbol{r_a = \frac{ln\left[\frac{2-\frac{2}{3}(0.12)}{0.123(0.12)}\right] ln\left[\frac{2-\frac{2}{3}(0.12)}{(0.1)0.123(0.12)}\right]}{(0.41)^2 u_2} \approx \frac{208}{u_2}} \tag{4.9}$$

The bulk surface resistance, r_s, [sm^{-1}] describes the resistance of vapour flow through the transpiring crop and evaporating soil surface:

$$\boldsymbol{r_s = \frac{r_l}{LAI_{active}}} \tag{4.10}$$

where r_l is bulk stomatal resistance of the well-illuminated leaf [sm^{-1}] and LAI_{active} is the active (sunlit) leaf area index [m^2m^{-2}]. LAI varies throughout the season and by crop type reaching a maximum before or at flowering. A generalised form of the above equation is:

$$\boldsymbol{r_s = \frac{r_l}{0.5\, LAI}} \tag{4.11}$$

where the LAI value is generalised as 24h [m^2m^{-2}] and the stomatal resistance, r_l, of a single leaf has a value of about 100 sm^{-1} under well-watered conditions. Hence:

$$r_s = \frac{100}{0.5\ (24 \times 0.12)} \approx 70 sm^{-1} \qquad (4.12)$$

4.4.4.2 Calculation of specific heat of air (c_p) and mean air density (ρ_a)

The specific heat of air (c_p) can be calculated by:

$$c_p = \frac{\gamma \varepsilon \lambda}{P} \qquad (4.13)$$

where c_p is specific heat at constant pressure [$MJkg^{-1}°C^{-1}$], γ is the psychometric constant [$kPa°C^{-1}$], ε is the ratio of molecular weight of vapour to dry air [0.622], λ is the latent heat of vaporisation [2.45 MJ kg^{-1} at 20°C] and P is the atmospheric pressure [kPa].

The mean air density (ρ_a) is calculated by:

$$\rho_a = \frac{P}{T_{Kv} R} \ where\ T_{Kv} = 1.01(T + 273) \qquad (4.14)$$

where R is the specific gas constant [0.287 $kJkg^{-1}K^{-1}$], T is the air temperature [°C] and T_{kv} is the virtual temperature [˚C].

Therefore:

$$\frac{c_p \rho_a}{r_a} = \frac{\gamma \varepsilon \lambda P}{P(1.01)(T + 273) R r_a} = \frac{\gamma (0.622) \lambda u_2}{(1.01)(T + 273)(0.287)(208)} \qquad (4.15)$$

Note that for daily evapotranspiration calculations the temperature T is set to:

$$T = \frac{T_{daily\ max} + T_{daily\ min}}{2} \qquad (4.16)$$

4.4.4.3 Penman-Monteith form of the reference crop evapotranspiration equation

Substituting the reference crop values of ρ_a, c_p, r_s and r_a (multiplied by 86,400 for units of day m^{-1}) into the combination equation and dividing through by λ to convert evapotranspiration to mm day^{-1}, assuming G to be nearly 0 for daily evapotranspiration calculations, yields:

$$ET_o = \frac{\Delta (R_n - G) + \frac{\gamma (0.622) \lambda u_2 \times 86,400}{(1.01)(T + 273)(0.287)(208)} (e_s - e_a)}{\Delta + \gamma \left(1 + \frac{70}{208} u_2\right)} \frac{u_2}{\lambda} \qquad (4.17)$$

$$ET_o \approx \frac{0.408\Delta(R_n) + \gamma \frac{900}{(T+273)} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)}$$

T can be calculated from the daily maximum and minimum temperatures developed in section 4.3 from MERRA data. Wind-speed at 2m height (u_2) is also available from MERRA data as northern and eastern wind-speed magnitudes and the resultant wind-speed magnitude calculated by:

$$u_2 = \sqrt{u_{north}^2 + u_{east}^2} \text{ [ms}^{-1}\text{]} \quad (4.18)$$

Figures 4.26 shows the MERRA maximum and minimum mean wind-speed magnitudes across Asia in the year 2000 (January 5th and October 4th).

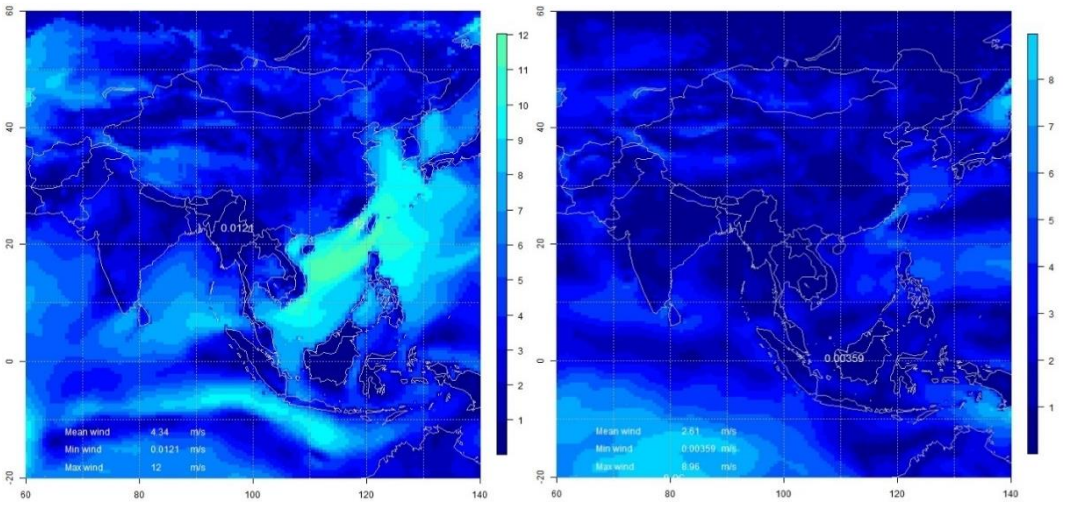


Figure 4.26: MERRA maximum (left, January 5th) and minimum (right, October 4th) mean daily wind-speed [ms^{-1}] across Asia for the year 2000 (based on data from Rienecker et al., 2011)

4.4.4.4 Calculation of slope of saturation vapour pressure temperature relationship (Δ)

The slope of the relationship between saturation vapour pressure and temperature, Δ, at a given temperature is given by:

$$\Delta = \frac{4098\left[0.6108exp\left(\frac{17.27T}{T+237.3}\right)\right]}{(T+237.3)^2} \quad (4.19)$$

where Δ is the slope of saturation vapour pressure curve at air temperature T [kPa °C^{-1}], T is the air temperature [°C] where T is the mean of the daily maximum and minimum temperature and exp[..] is 2.7183 (base of natural logarithm) raised to the power [..]. Therefore, calculation of Δ only requires MERRA daily maximum and minimum temperatures.

4.4.4.5 Calculation of net radiation (R_n)

R_n is energy available per unit area, net raditation, and expressed in MJ m^{-2} day^{-1}. To convert the energy units for radiation to equivalent water depths [mm], the latent heat of vaporization, λ, is used as a conversion factor.

$$\boldsymbol{R_n = R_{ns} - R_{nl}} \quad (4.20)$$

where R_{ns} is the net incoming shortwave radiation at the surface [MJm^{-2}day^{-1}] and R_{nl} is the net outgoing longwave radiation at the top of the atmosphere (TOA) [MJm^{-2}day^{-1}].

The incoming net shortwave radiation at the surface R_{ns} can be represented by the fraction of net surface downward shortwave radiation flux (R_s) [MJm^{-2}day^{-1}] that is not reflected from the surface:

$$\boldsymbol{R_{ns} = (1 - \alpha)R_s} \quad (4.21)$$

where α is the surface albedo [fraction].

Albedo is the proportion of incident light or radiation that is reflected by the Earth's surface, usually represented as a value between 0 (i.e. all the incident radiation is absorbed) or 1 (i.e. all the incident radiation is reflected). Dark surfaces usually have low albedo values where as bright surfaces (e.g. sand, snow) have high albedo values.

The difference between outgoing and incoming longwave radiation is called the net longwave radiation, R_{nl}. As the outgoing longwave radiation is almost always greater than the incoming longwave radiation, R_{nl} represents an energy loss.

$$\boldsymbol{R_{nl} = R_{emittedl} - R_{absorbedl}} \quad (4.22)$$

where $R_{emittedl}$ is the emitted longwave radiation at the surface [MJm^{-2}day^{-1}] and $R_{absorbedl}$ is the absorbed longwave radiation at the surface [MJm^{-2}day^{-1}].

Therefore, net radiation is calculated by:

$$R_n = (1 - \alpha)R_s - R_{emittedl} - R_{absorbedl} \quad (4.23)$$

R_n, α, $R_{emittedl}$ and $R_{absorbedl}$ are all available within the MERRA data suite. Unsurprisingly, incoming solar radiation and absorbed and emitted longwave radiation are at a minimum in winter (December/January) and at a maximum during the spring/summer months (April to August). Monsoon conditions (cloud etc) disrupts radiation and therefore peak incoming solar radiation is often seen in April/May. Figures 4.27, 4.28 and 4.29 show minimum and maximum daily incoming solar, absorbed and emitted longwave radiation (respectively) for the year 2000.

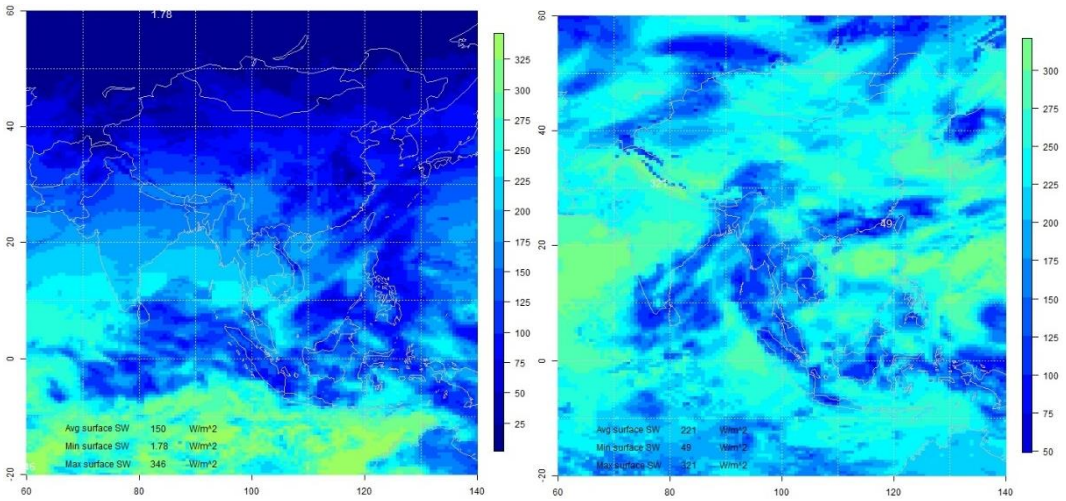


Figure 4.27: MERRA minimum (left, December 17th) and maximum (right, May 2nd) mean net surface downward shortwave radiation flux [Wm-2] across Asia for the year 2000 (based on data from Rienecker et al., 2011)

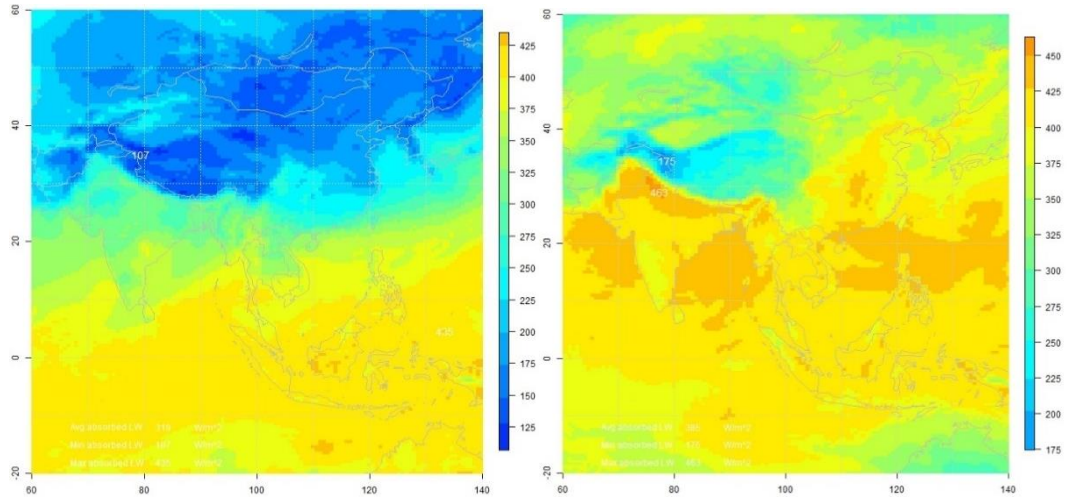


Figure 4.28: MERRA minimum (left, January 26th) and maximum (right, July 6th) mean absorbed longwave radiation [Wm-2] across Asia for the year 2000 (based on data from Rienecker et al., 2011)

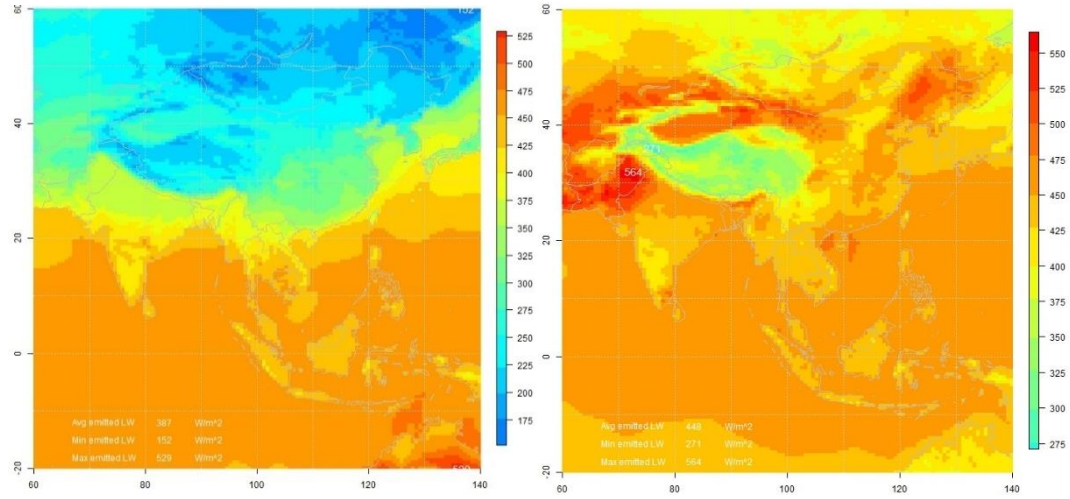


Figure 4.29: MERRA minimum (left, January 17th) and maximum (right, July 9th) mean emitted longwave radiation [Wm^{-2}] across Asia for the year 2000 (based on data from Rienecker et al., 2011)

Albedo shows a reverse trend with maximum albedo experienced in winter months when snow coverage is at a maximum and vegetation at a minimum, and minimum albedo values seen in later summer (August/September) as snow levels are minimal and vegetation is at full growth. Figure 4.30 shows minimum and maximum albedo for the year 2000.

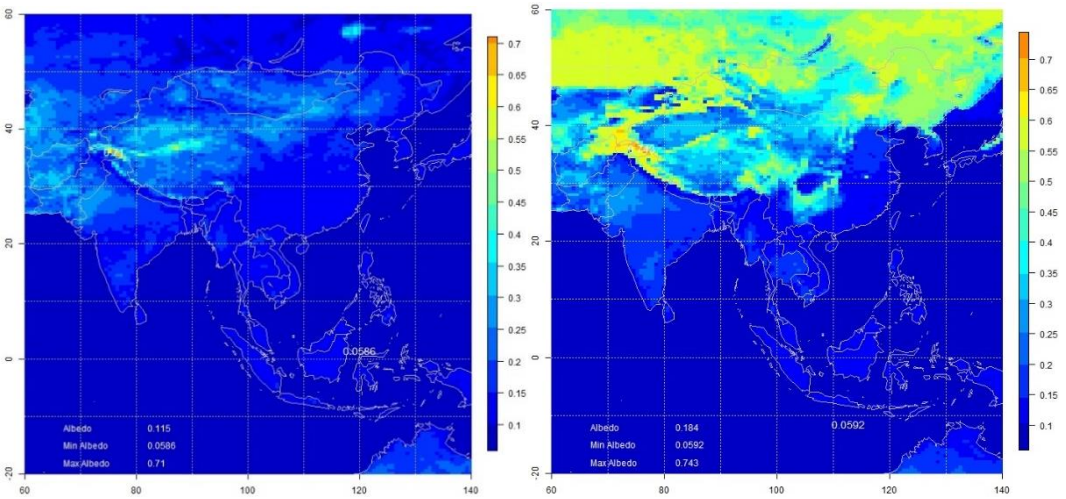


Figure 4.30: MERRA minimum (left, September 24th) and maximum (right, January 31st) mean albedo [dimensionless] across Asia for the year 2000 (based on data from Rienecker et al., 2011)

4.4.4.6 Calculation of psychrometric constant (γ)

The psychrometric constant, γ [kPa°C^{-1}], relates the partial pressure of water in air to the air temperature and is given by:

$$\gamma = \frac{c_p P}{\varepsilon \lambda} = 6.65 * 10^{-4} P \qquad (4.24)$$

where P is the atmospheric pressure [kPa], λ is the latent heat of vaporization [2.45 MJkg^{-1}], c_p is the specific heat at constant pressure, [1.013x10^{-3} MJ kg^{-1}°C^{-1}] and ε is the ratio of molecular weight of water vapour/dry air [0.622].

Pressure data is also available from MERRA. Although there are seasonal variations in air pressure, elevation is a bigger driver of air pressure change and hence the Tibet/Qinghai plateau has very low mean air pressure values (see figure 4.31).

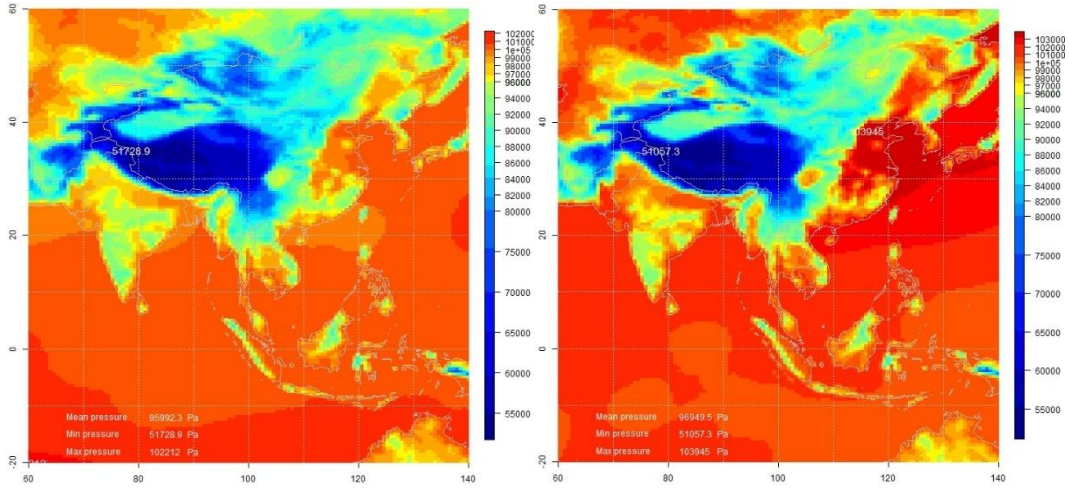


Figure 4.31: MERRA minimum (left, July 8th) and maximum (right, January 29th) mean air pressure [Pa] across Asia for the year 2000 (based on data from Rienecker et al., 2011)

4.4.4.7 Calculation of vapour pressure deficit ($e_s - e_a$)

The vapour pressure deficit is the difference between the saturation (e_s) and actual vapour pressure (e_a) for a given time-period, and is related to air temperature by:

$$e^o(T) = \left(0.6108 exp\left[\frac{17.27T}{T + 237.3}\right]\right) \quad (4.25)$$

where e°(T) is the saturation vapour pressure at the air temperature T [kPa], T is the air temperature [°C] and exp [..] is the 2.7183 (base of natural logarithm) raised to the power [..].

The mean saturation pressure is calculated by:

$$e_s = \frac{e^o(T_{max}) + e^o(T_{min})}{2} \quad (4.26)$$

$$e_s = 0.3054\left(exp\left[\frac{17.27T_{max}}{T_{max} + 237.3}\right] + exp\left[\frac{17.27T_{min}}{T_{min} + 237.3}\right]\right)$$

The actual vapour pressure can also be calculated from the relative humidity:

$$e_a = \frac{e^o(T_{min})\frac{RH_{max}}{100} + e^o(T_{max})\frac{RH_{min}}{100}}{2} \quad (4.27)$$

$$e_a = 0.3054\left(exp\left[\frac{17.27T_{min}}{T_{min} + 237.3}\right]\frac{RH_{max}}{100} + exp\left[\frac{17.27T_{max}}{T_{max} + 237.3}\right]\frac{RH_{min}}{100}\right)$$

where T_{min} is the daily minimum temperature [°C], T_{max} is the daily maximum temperature [°C], RH_{max} is the daily maximum relative humidity [%] and RH_{min} is the daily minimum relative humidity [%].

Relative humidity (RH) is not available from the MERRA dataset but can be calculated from the specific humidity (SH) [kg kg^{-1}] which is available (Cactus, 2001; Lowe and Ficke, 1974). SH can be expressed as:

$$\boldsymbol{SH = \frac{x_{H_2O}.M_{H_2O}}{x_{H_2O}.M_{H_2O} + (1 - x_{H_2O}).M_{dry}}} \quad (4.28)$$

where x_{H2O} = mole fraction, volume mixing ratio of water, M_{H2O} = molar mass of water [0.0181534 $kgmol^{-1}$] and M_{dry} is the molar mass of dry air [0.0289644 $kgmol^{-1}$].

The value of x_{H2O} can be calculated by:

$$\boldsymbol{x_{H_2O} = \frac{P_{H_2O}}{P}} \quad (4.29)$$

where P_{H20} is the partial pressure of water [Pa] and P is the air pressure [Pa].

Relative humidity, RH [%], can be calculated by:

$$\boldsymbol{RH = \frac{e(T_D)}{e(T)}.100 = \frac{P_{H_2O}}{e(T)}.100} \quad (4.30)$$

where $e(T_D)$ is the vapour pressure at the dew point [Pa] and e(T) is the vapour pressure at the air temperature [Pa].

Hence:

$$\boldsymbol{P_{H_2O} = \frac{RH.e(T)}{100}} \quad (4.31)$$

$$\boldsymbol{x_{H_2O} = \frac{RH.e(T)}{100P}}$$

$$\boldsymbol{SH = \frac{\frac{RH.e(T)}{100P}.M_{H_2O}}{\frac{RH.e(T)}{100P}.M_{H_2O} + \left(1 - \frac{RH.e(T)}{100P}\right).M_{dry}}}$$

$$\boldsymbol{RH = \frac{100P.M_{dry}}{e(T).(\frac{M_{H_2O}}{SH} - M_{H_2O} + M_{dry})}}$$

From Lowe and Ficke (1974), the value of e(T) can be calculated by:

$$e(T) = a_0+T\cdot(a_1+T\cdot(a_2+T\cdot(a_3+T\cdot(a_4+T\cdot(a_5+T\cdot a_6)))))$$

$$\mathbf{e(T) = \min(e_{water}, e_{ice}) \text{ ,-50° C} \leq \mathbf{T} \leq \mathbf{100 \text{ °C}}} \qquad (4.32)$$

Parameter	Water	Ice
a_0	6.107799961	6.109177956
a_1	$4.436518521 \times 10^{-1}$	$5.034698970 \times 10^{-1}$
a_2	$1.428945805 \times 10^{-2}$	$1.886013408 \times 10^{-2}$
a_3	$2.650648471 \times 10^{-4}$	$4.176223716 \times 10^{-4}$
a_4	$3.031240396 \times 10^{-6}$	$5.824720280 \times 10^{-6}$
a_5	$2.034080948 \times 10^{-8}$	$4.838803174 \times 10^{-8}$
a_6	$6.136820929 \times 10^{-11}$	$1.838826904 \times 10^{-10}$

Table 4.4: Parameters for calculation of saturation vapour pressure for water and ice (Lowe and Ficke, 1974)

Temperature, pressure and specific humidity data are available from MERRA. Specific humidity varies throughout each day and follows a seasonal pattern with low values in winter and high humidity in summer (see figure 4.32). Figure 4.33 shows relative humidity plotted on the same days as in figure 4.32.

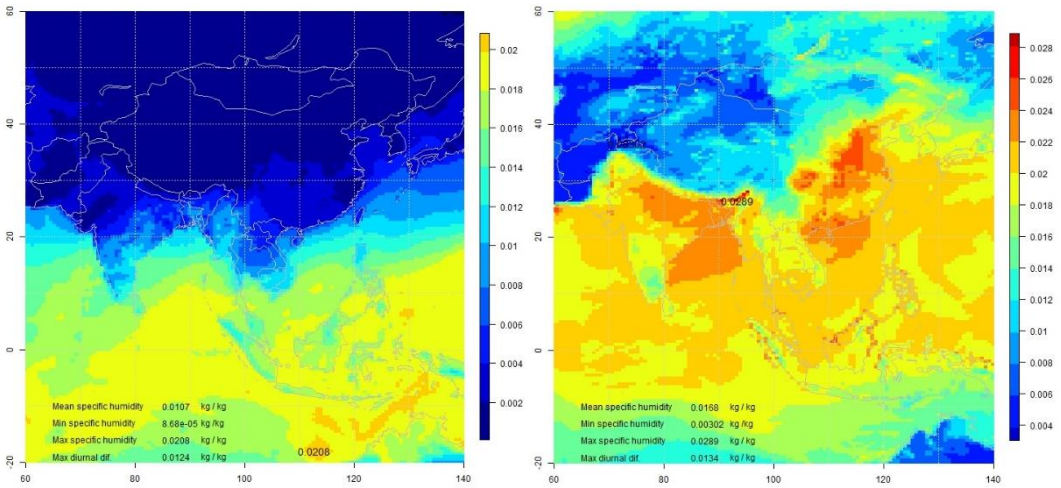


Figure 4.32: MERRA daily minimum (left, January 27th) and maximum (right, July 17th) mean specific humidity [kg kg^{-1}] across Asia for the year 2000 (based on data from Rienecker et al., 2011)

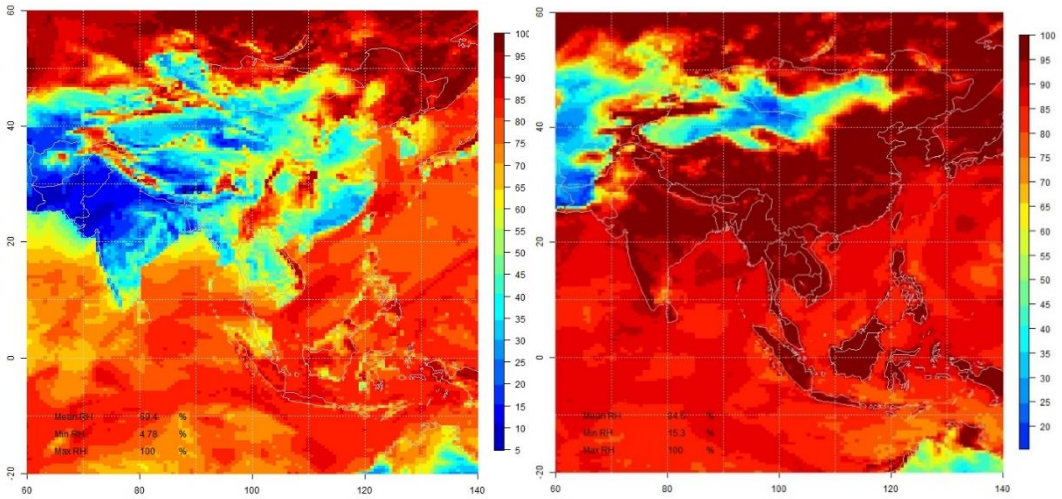


Figure 4.33: Relative humidity daily minimum (left, January 27th) and maximum (right, July 17th) [%] across Asia for the year 2000 calculated using the Cactus methodology (based on equations within Cactus, 2001; Lowe and Ficke, 1974; and data from Rienecker et al., 2011)

4.4.5 Preparation of the daily evapotranspiration datasets

Daily reference evapotranspiration datasets for Asia were developed by the author using the FAO method using MERRA datasets as input from 1979 to 2013 (although only 1979 to 2007 was used in the hydrological model). The datasets were cropped to the same dimensions as the precipitation and temperature datasets. For compatibility, the evapotranspiration datasets were increased in resolution to that of the APHRODITE datasets (i.e. 0.25° x 0.25°) using the R 'disaggregate' function. In disaggregation, the values in the new higher resolution grid are the same as in the larger underlying original cells.

4.4.6 Characteristics of the developed Yangtze basin evapotranspiration datasets

The Asia datasets as developed in 4.4.4 show low winter evapotranspiration and high summer evapotranspiration (see figures 4.34 and 4.35).

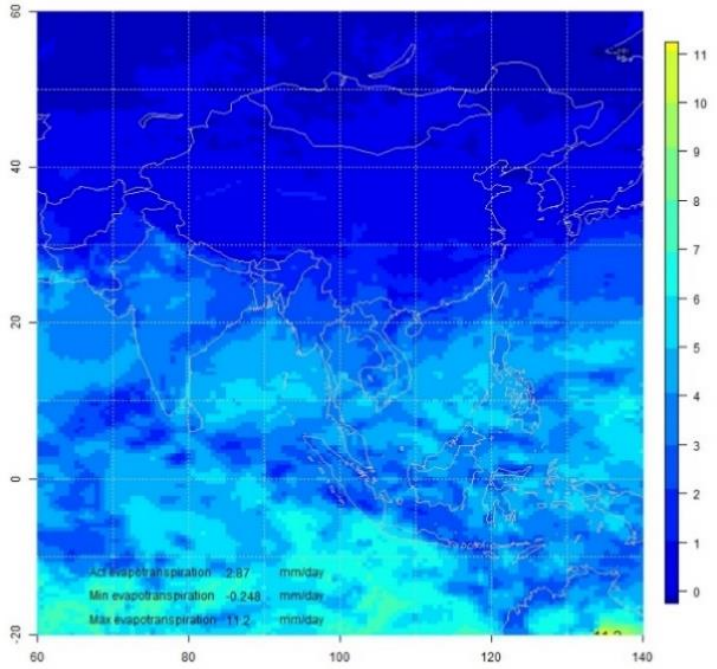


Figure 4.34: Minimum mean evapotranspiration [mm day^{-1}] across Asia for the year 2000 (January 12th) developed using the FAO method and MERRA datasets

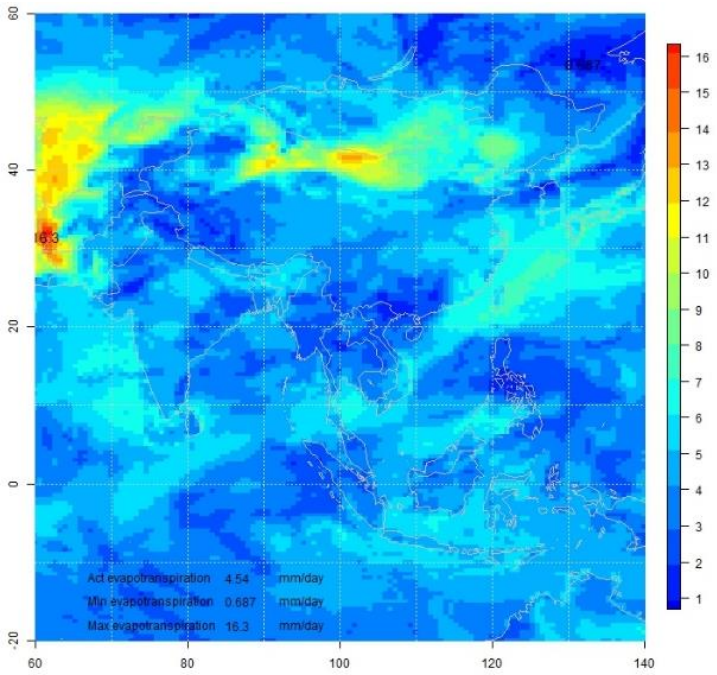


Figure 4.35: Maximum mean evapotranspiration [mm day^{-1}] across Asia for the year 2000 (July 22nd) developed using the FAO method and MERRA datasets

When cropped to the YDB, seasonal variation in evapotranspiration reveals low rates in winter and high rates in summer, with lowest values in the high mountains and plateau to the north and west (see figures 4.36, 4.37, 4.38, 4.39, 4.40).

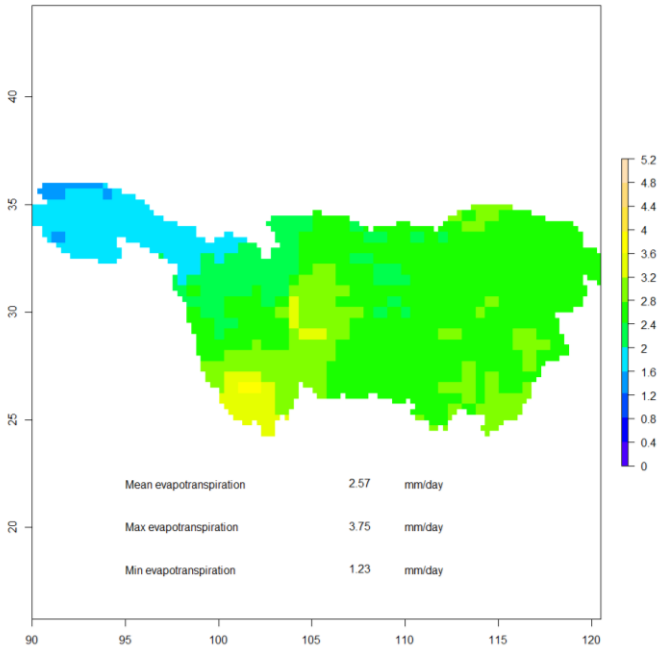


Figure 4.36: Mean annual evapotranspiration [mm day^{-1}] 1979-2013 across the YDB developed using the FAO method and MERRA datasets

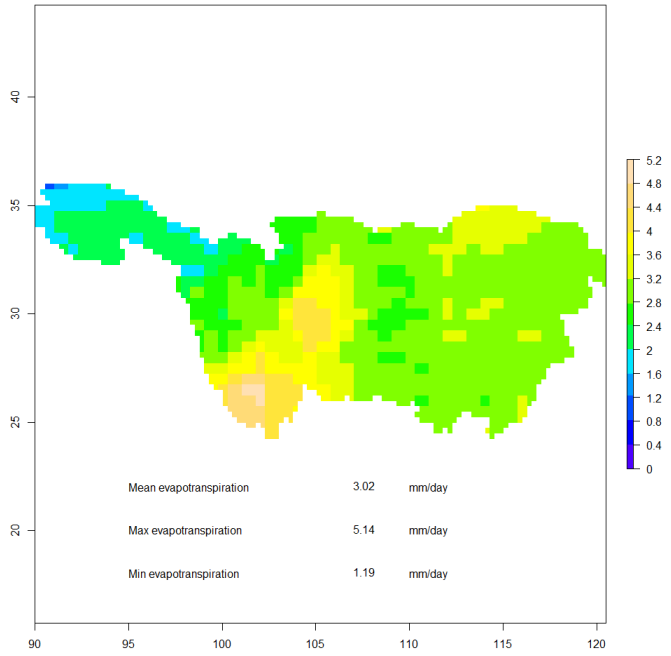


Figure 4.37: Mean spring (Mar, Apr, May) evapotranspiration [mm day^{-1}] 1979-2013 across the YDB developed using the FAO method and MERRA datasets

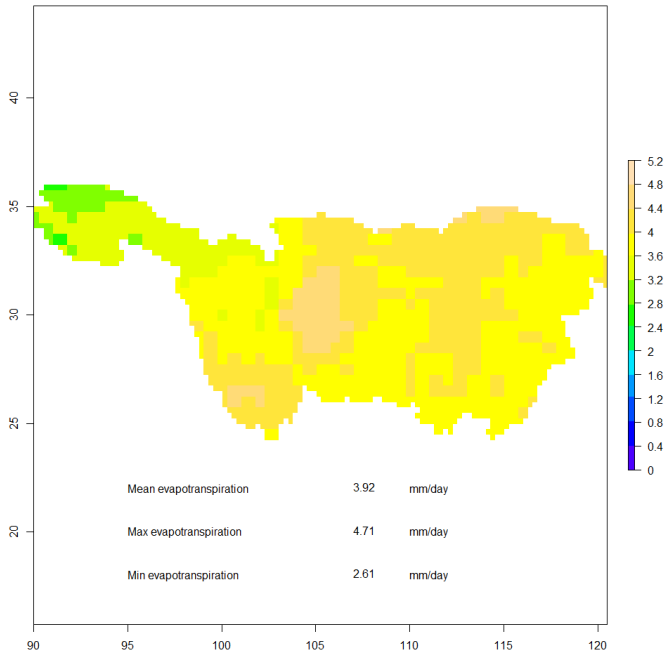


Figure 4.38: Mean summer (Jun, Jul, Aug) evapotranspiration [mm day^{-1}] 1979-2013 across the YDB developed using the FAO method and MERRA datasets

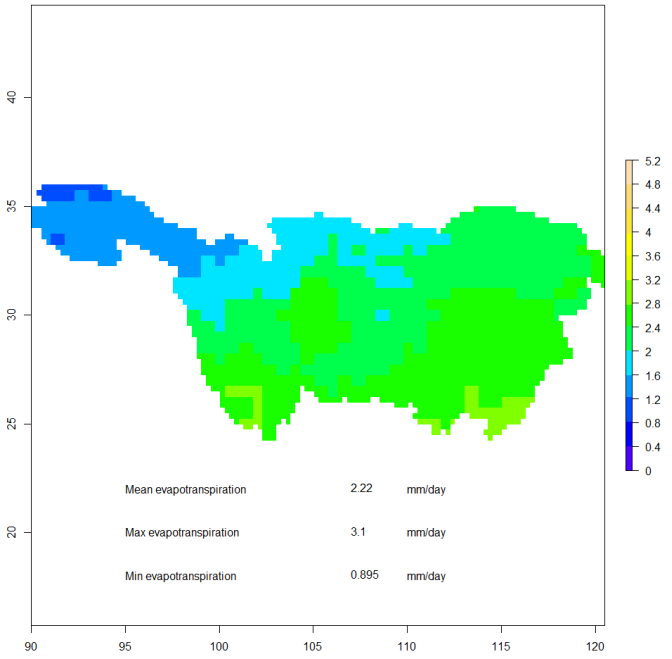


Figure 4.39: Mean autumn (Sep, Oct, Nov) evapotranspiration [mm day^{-1}] 1979-2013 across the YDB developed using the FAO method and MERRA datasets

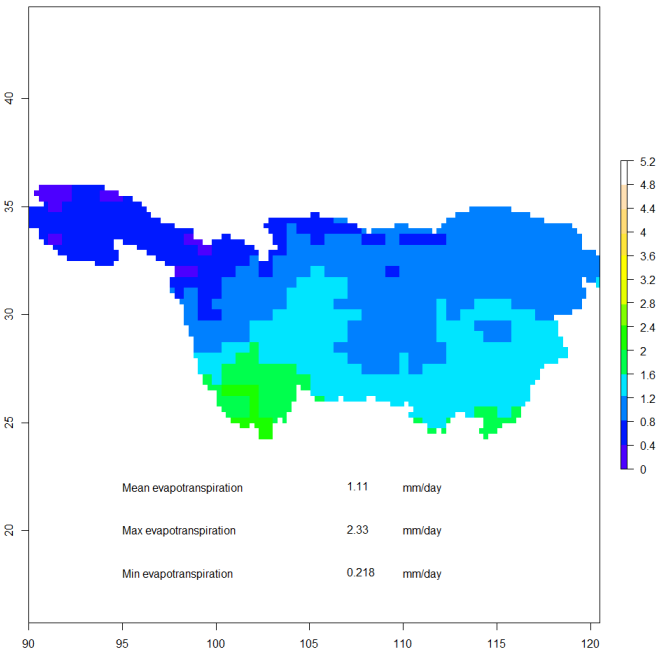


Figure 4.40: Mean winter (Dec, Jan, Feb) evapotranspiration [mm day^{-1}] 1979-2013 across the YDB developed using the FAO method and MERRA datasets

4.4.7 Annual variation in the evapotranspiration dataset

As temperatures are increasing across the YDB (see figure 4.23) it is unsurprising that running mean annual reference crop evapotranspiration is also rising over the record of the dataset (see figure 4.41). Higher evapotranspiration rates coupled with steady or reducing precipitation (see figure 4.10) could result in lower surface discharge and hence reduced hydropower potential.

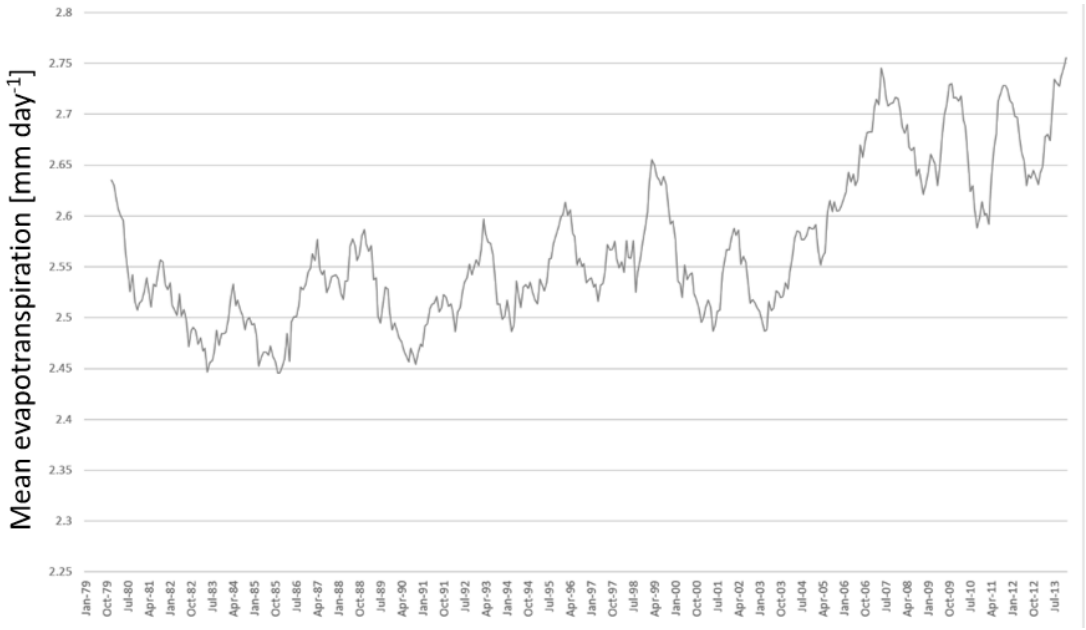


Figure 4.41: Running mean annual reference crop evapotranspiration [mm day^{-1}] 1979-2013 across the YDB developed using the FAO method and MERRA datasets

4.4.8 Validation of the evapotranspiration dataset

The evapotranspiration dataset developed was compared to the University of East Anglia Climatic Research Unit (CRU) high-resolution potential evapotranspiration (pet) dataset (Harris et al., 2014). The 'pet' gridded dataset within CRU TS v3.24 has a monthly temporal resolution and 0.5° x0.5° spatial resolution. The developed daily data for each month was combined into monthly rasters and the overall mean of both the developed and CRU data compared between 1981 and 1990. The root mean square error (RMSE) is 0.37mm day^{-1}. Figure 4.42 shows how the ratio between the developed dataset and the CRU dataset varies month by month and there is a seasonal pattern, with the ratio peaking in May/June and is minimal in winter. Figure 4.43 shows the spatial difference between the annual mean of the developed datasets and the CRU dataset. Annual mean error is between -0.7 and 1.2 mm

day^{-1}, although most of the error greater than 0.5mm is concentrated in the mid-Yangtze (a major rice growing area).

It is hypothesised that this error is largely explained due to the difference in calculating reference crop evapotranspiration and the fact that the CRU data is potential evapotranspiration, the latter of which is a product of reference evapotranspiration and crop coefficients (see equation 4.1). Crop coefficients vary as crops develop, ranging from 0.3 to 1.0 at the beginning of the growing cycle and peaking between 0.7 to 1.25 in mid growing season (Allen et al., 1998). April and May are at the beginning of the growing cycle of many of the Yangtze's main crops, including rice, maize and wheat. The ratio range of 0.8 and 1.2 in figure 4.43 is within the range of changing crop coefficients. Although the MERRA dataset is attractive due to its daily resolution, the CRU dataset is also attractive for future hydropower search elsewhere due to the inclusion of crop coefficients, higher spatial resolution and would be less arduous to prepare for other catchments.

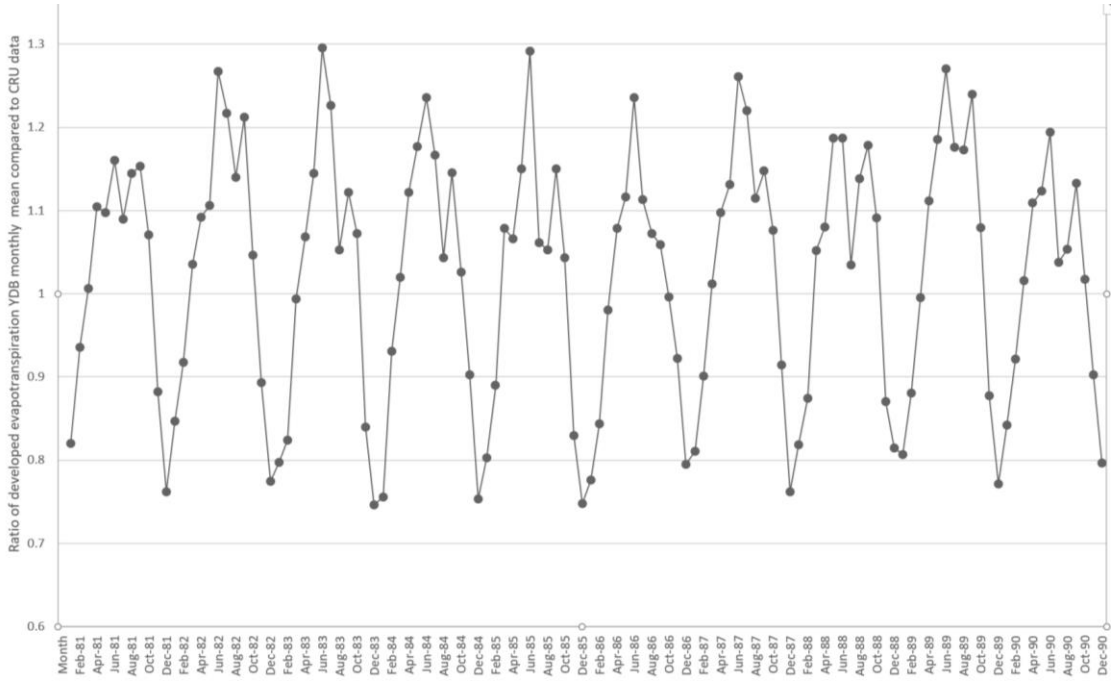


Figure 4.42: Ratio of developed evapotranspiration dataset monthly mean across the YDB to CRU monthly mean data, 1979-2013

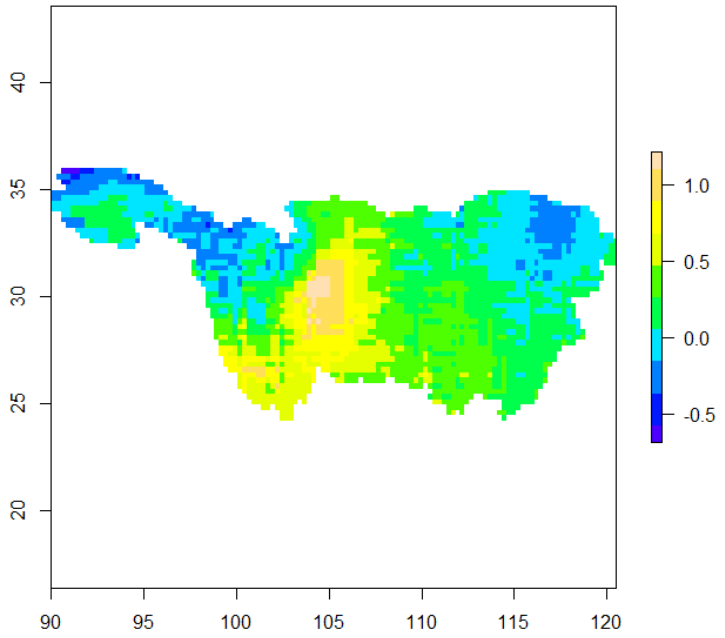


Figure 4.43: Mean annual error in evapotranspiration [mm day^{-1}] between the developed MERRA dataset and CRU datasets

4.5 Development of a snowfall/accumulation/melt model

4.5.1 An introduction to snowfall/accumulation/melt models

Snowfall accumulation and melt requires consideration when hydrological modelling of drainage basins at high altitudes or latitudes. The meteorological datasets prepared thus far have all been available or developed as stand-alone gridded datasets. Although stand-alone gridded datasets do exist for snow depth and coverage, the snowfall/accumulation/melt model required here needs to maintain a water balance with the precipitation grids developed earlier. Hence, a model manipulating the incoming precipitation based on temperature is preferred.

Air temperature is the most commonly used parameter in separating precipitation into rain and snow within snowmelt models, and is highly correlated to radiation, wind and humidity which are important components of heat transfer within a snowpack. Since air temperature is relatively easy to measure and existing temperature datasets have wide coverage, it is natural that snow models are often represented as a function of temperature. It is also relatively straight-forward to estimate temperature in ungauged areas as variability is determined primarily through elevation, an important point in snow modelling as most snow processes

occur at heights above measurement stations. Daily melt models are often represented by simple degree-day models where melt is predicted depending on the difference between air temperature and a reference melt temperature (usually near to 0°C), multiplied by a degree-day factor:

$$\boldsymbol{W_{melt} = D \times (T_{air} - T_{melt})} \quad (4.33)$$

where W_{melt} is snowmelt [mm], D is a degree day factor [$mm°C^{-1}day^{-1}$], T_{air} is air temperature [°C] and T_{melt} is the melt temperature [°C].

Although simple, and computationally inexpensive to calculate, degree-day factors are difficult to estimate and require calibration as they vary due to location and season. More sophisticated energy balance models are available (e.g. Marks et al., 1999; Tarboton et al., 1994; Hock and Holmgren, 2005) taking temperature, radiation, humidity, precipitation and wind speed as input. Although a better representation of snow dynamics, this comes at a computing cost.

In this project, a compromise was sought by adopting a simple temperature-based separation of precipitation into snow and rain components, and a more sophisticated conceptual index melt model.

4.5.2 The Snow Accumulation and Ablation Model – SNOW 17

In a review of models and procedures for modelling urban snowmelt (Moghadas et al., 2016), only two models were classified as fully distributed (i.e. snowmelt can be estimated grid by grid), a necessary requirement to integrate into the gridded datasets developed thus far. One of these was the National Weather Service River Forecast System (NWSRFS) Snow Accumulation and Ablation Model, referred to as SNOW-17 (Anderson, 1973; 2006). In an evaluation of thirty-three snowpack models across a wide range of hydrometeorological and forest canopy conditions at five Northern Hemisphere sites (Rutter et al., 2009), SNOW-17 snow water equivalent (SWE) measurements compared favourably with on-site observations with a mean normalised RMSE of <1.5mm in open areas and <1mm in forested areas. SNOW-17 is a conceptual model primarily used in river forecasting, employing air temperature as the sole index to determine energy exchange across the snow-air interface, making it a particularly attractive choice of model.

A comparison between an earlier version of SNOW-17 and a detailed energy balance model showed SNOW-17 compared well, except during:

1. Periods with warm temperatures, high humidity, and strong winds
2. Clear sky periods with an aged snow surface (i.e. low albedo) and cold temperatures
3. Periods with air temperatures well above normal but calm conditions

4.5.3 A description of the SNOW-17 model

This section describes the SNOW-17 model as developed by Anderson, 1973; 2006. SNOW-17 includes most of the important physical processes that take place within snow cover but only in a simplified form, including:

- Form of precipitation
- Accumulation of the snow cover
- Energy exchange at the snow-air interface
- Internal state of the snow cover
- Transmission of water through the snow cover
- Heat transfer at the snow-soil interface

Sections 4.5.3.1 to 4.5.3.9 describe the philosophy and equations within the SNOW-17 proprietary model that were included within the hydrological modelling software developed by the author of this work.

4.5.3.1 Form of precipitation

Surface air temperature is a good, but not perfect, indicator of the form of precipitation. Data from Snow Investigations (Snow Hydrology, 1956) shows how precipitation can fall as rain at surface air temperatures below -1°C and as snow at temperatures above 4°C. However, the typical temperature separating rain and snow is 1.5°C, and in this project all precipitation falls as rain above T_{rain} (1.5°C) and all precipitation falls as snow below T_{snow} (0°C). In between these temperatures, there is a mix of rain and snow and the fraction of snow is determined by:

$$f_{snow} = \frac{\overline{T} - T_{snow}}{T_{rain} - T_{snow}} \qquad (4.34)$$

where f_{snow} is the fraction of snow in the mixed rain/snow, $\overline{T}$ is the mean daily temperature [°C], T_{rain} is the threshold rain temperature [1.5 °C] and T_{snow} is the threshold snow temperature

[0 °C]. In hydrological modelling, it is the SWE that is of interest rather than the depth of the snow.

4.5.3.2 Accumulation of the snow cover

Snow can be lost to sublimation before settling on the ground, which occurs when the vapour pressure of the air is less than the vapour pressure at the snow surface. However, in this model all falling precipitation from the APHRODITE data is partitioned entirely into snow and rain without losses. The temperature of new snow is set to:

$$\boldsymbol{T_{new} = min(0, T_{air})} \tag{4.35}$$

where T_{new} is the temperature of new snow [°C] and T_{air} is the air temperature [°C], and the addition of new snowfall increases the heat deficit (the amount of heat required to be added to new snowfall to bring it up to 0°C) by:

$$\boldsymbol{\Delta D_p = -\frac{c_i(T_{new} \cdot S_{new})}{L_f}} \tag{4.36}$$

where ΔD_p is the change in the heat deficit due to snowfall [mm], T_{new} is the temperature of new snow [°C], S_{new} is the new snowfall amount [mm], c_i is the specific heat of ice [0.5 cal gm^{-1} $°C^{-1}$] and L_f is the latent heat of fusion [80 cal gm^{-1}].

4.5.3.3 Energy exchange at the snow-air interface

Energy exchange between the air and snow is a factor of net radiation, latent and sensible heat transfer and heat within precipitation. Due to the high albedo of snow, much of the incoming solar radiation is reflected to space, and fresh snow has a much higher albedo (approximately 90%) compared to aged snow (approximately 40%). Shortwave radiation is also affected by vegetation and cloud cover, slope and aspect. Snow is a very efficient producer of longwave radiation with net longwave radiation positive when the air is warm and the sky overcast, and negative when the skies are clear and temperatures cool.

Latent and sensible heat transfers are turbulent processes and dependent on wind speed, and the latent heat exchange is a function of the vapour pressure gradient which can be calculated from the dew-point temperature. Vapour pressure at the snow surface is equal to the saturation vapour pressure at the snow surface temperature and when the vapour pressure of the air is

greater than that at the snow surface, vapour is transferred from the air to the snow, releasing heat. The reverse takes place when the vapour pressure of the air is lower.

Sublimation most commonly occurs during accumulation periods which requires heat to convert the solid to vapour, whilst condensation dominates melt periods. Latent heat transfer produces a small net overall sublimation loss, particularly when humidity is low and windy conditions persist. Sensible heat exchange is a function of the temperature gradient between the air and snow surface, with positive exchange when the air is warmer than the snow. As rain enters snow, heat is transferred to the rain water until the water temperature comes into equilibrium with the snow, a function of the amount of precipitation and temperature of the rain.

When the net heat exchange is positive, and the snow surface is at 0°C, the snow melts as the temperature of the snow cannot rise any further. Below 0°C, heat exchange can be positive or negative, dependent on the temperature gradient in the upper layers, and the rate of heat transfer is low due to snow being a good insulator, a function of the thermal conductivity which is largely dependent on density. Density increases as the melt season progresses and the snow 'ripens'.

The model keeps track of the heat deficit and expresses the energy exchange in terms of mm, where mm of energy is the amount of heat required to melt or freeze 1mm of ice or water – approximately 8 cal/cm^2, making it easy to compare heat deficit with an amount of melt or rain water to overcome the deficit.

4.5.3.4 Rain-on-snow surface melt computations

When rain falls onto snow:

- incoming solar radiation is negligible due to overcast conditions
- incoming longwave radiation is equal to black body radiation (i.e. an emissivity of 1.0) at the air temperature
- relative humidity is high (90%)
- the surface temperature of the snow is 0°C

If the amount of rain in one hour is greater than 0.25mm, the rain-on-snow melt can be calculated by:

$$M_r = \sigma \Delta t_p [(T_a + 273)^4 - 273^4] + 0.0125 P_r T_r + 8.5 UADJ \frac{\Delta t_p}{6} [(0.9 e_{sat} - 6.11) + 0.00057 P_a T_a] \quad (4.37)$$

where M_r is melt during rain-on-snow time intervals [mm], σ is the Stefan-Boltzman constant [6.12×10^{-10} mm˚K^{-1}hr^{-1}], Δt_p is the time interval of precipitation data [hours], T_a is the air temperature [˚C] where 273K is equal to 0˚C on the Kelvin scale, P_r is the precipitation in the form of rain [mm] and T_r is the temperature of rain [˚C], equal to T_a or 0˚C (whichever is greater). UADJ is a model parameter representing the average wind function that is adjusted during model calibration and is site dependent. Here it is set to 0.04 [mm mb^{-1} 6hr^{-1}], at the lower end of suggested initial parameter values. P_a is the atmospheric pressure [mb], H_e is height elevation [m] and e_{sat} is the saturated vapour pressure at T_a [mb] – computed from:

$$e_{sat} = 2.7489 \times 10^8 e^{\left(-\frac{4278.63}{T_a + 242.792}\right)} \quad (4.38)$$

4.5.3.5 Non-rain surface melt computations

When there is no rain or rain is light (<0.25mm) and the air temperature is above a base value, the melt is calculated as:

$$M_{nr} = M_f (T_a - MBASE) + 0.0125 P_r T_r \quad (4.39)$$

where M_{nr} is melt during non-rain periods [mm], M_f is a melt factor [mm°C$^{-1}\Delta t^{-1}$] and MBASE is a model parameter representing base temperature [here set as 0 °C].

During non-rain periods, melt is significantly affected by solar radiation which varies throughout the year, and in the northern hemisphere is at a maximum around June 21st and at a minimum around December 21st. Albedo also has a seasonal variation as fresh snow is more common in mid-winter and aged snow more common as the melt season progresses. Hence the melt factor (M_f) should vary seasonally and expressed as:

$$M_f = \frac{\Delta t}{6} [S_v A_v (MFMAX - MFMIN) + MFMIN] \quad (4.40)$$

$$S_v = 0.5 sin\left(\frac{2\pi N}{366}\right) + 0.5 \quad (4.41)$$

where A_v is seasonal variation adjustment (set to 1.0 for latitudes <54°N), MFMAX is a model parameter representing the maximum melt factor, assumed to be on June 21st [1.025

mm°C^{-1}6h^{-1}], MFMIN is also a model parameter representing the minimum melt factor assumed to be on December 21st [0.4 mm°C^{-1}6h^{-1}] and N is the number of days since March 21st.

4.5.3.6 Energy exchange without surface melt

SNOW-17 tracks the heat deficit which changes when the air temperature is below freezing due to energy exchange across the snow-air interface. When below freezing the thermal gradient in the upper layers of the pack influences whether the snow cover loses or gains heat. The thermal gradient is estimated as the difference between the snow surface temperature (T_{sur}) and the temperature some distance within the snow pack by computing the antecedent temperature index (ATI):

$$\boldsymbol{T_{sur} = min(0, T_a)} \tag{4.42}$$

$$\boldsymbol{ATI_2 = ATI_1 + TIPM_{\Delta t}(T_a - ATI_1)} \tag{4.43}$$

$$\boldsymbol{TIPM_{\Delta t} = 1.0 - (1.0 - TIPM)^{\Delta t/6}} \tag{4.44}$$

where T_{sur} is the snow surface temperature [°C], T_a is the air temperature [°C], ATI_2 is the antecedent temperature index at the current time step [°C], ATI_1 is the antecedent temperature index at the previous time step [°C] and TIPM is a model parameter, set to 0.1.

When T_{sur} is less than ATI, the heat deficit is increasing, and the reverse is true if T_{sur} is greater than ATI. The change in the heat deficit due to a temperature gradient in the surface layers of the snow cover is calculated by:

$$\boldsymbol{\Delta D_t = NMF \frac{\Delta t}{6} \frac{M_f}{MFMAX} (ATI - T_{sur})} \tag{4.45}$$

where ΔD_t is the change in heat deficit due to a temperature gradient [mm], NMF is a model parameter representing negative melt factor [0.15 mm°C^{-1}6hr^{-1}], Δt is the time-period [hours], M_f is a melt factor [mm°C$^{-1}\Delta t^{-1}$] and MFMAX is the maximum melt factor on June 21st [1.025 mm°C^{-1}6h^{-1}].

4.5.3.7 Internal state of the snow cover

Snow cover can contain water in liquid and solid form as liquid water adheres to ice crystals and in well-aged snow can be 2-10% by weight. As the snow cover loses heat as temperatures

drop below 0°C, any liquid water present in the surface layers will refreeze, and if temperatures persist, liquid water deeper into the pack will also refreeze. The movement of heat depends on the thermal conductivity which is largely dependent on density. Density also causes depth changes even if the SWE remains constant and is affected by snowfall, compaction, destructive/constructive and melt metamorphism.

SNOW-17 treats the snow cover as a single lumped entity and does not try to calculate the temperature, liquid water or density profile, but accounts for the ripeness of the snow cover by tracking the heat deficit and liquid water storage. A snow cover is ripe when any additional melt or rain water moves through the pack to become outflow and occurs at isothermal temperatures (0°C) (i.e. the heat deficit is zero) and the liquid water capacity is full, determined by:

$$\boldsymbol{W_{qx} = PLWHC \times W_i} \tag{4.46}$$

where W_{qx} is the liquid water capacity [mm], PLWHC is a model parameter representing the percent liquid water holding capacity [0.04] and W_i is the water equivalent of the ice portion of the snow cover [mm]. To account for the heat deficit and liquid water storage for each time interval, SNOW-17 first calculates the liquid water at the surface due to melt and rain:

$$\boldsymbol{Q_w = M_r + M_{nr} + P_r} \tag{4.47}$$

where Q_w is the liquid water available at the snow surface [mm], M_r is the melt during rain-on-snow time intervals [mm], M_{nr} is the melt during non-rain periods [mm] and P_r is the precipitation in the form of rain [mm]. The heat deficit is then adjusted to the temperature of the new snowfall and heat transfer caused by a temperature gradient in the upper layers is calculated by:

$$\boldsymbol{D_2 = D_1 + \Delta D_p + \Delta D_t} \tag{4.48}$$

where D_2 is the heat deficit at the current timestep [mm], D_2 is the heat deficit at the previous timestep [mm], ΔD_p is the change in the heat deficit due to snowfall [mm] and ΔD_t is the change in heat deficit due to a temperature gradient [mm]. If there is sufficient water available

at the surface to overcome the heat deficit and the liquid water storage capacity is exceeded, the snow cover is ripe, and any excess water becomes outflow:

$$\boldsymbol{E = Q_w + W_q - W_{qx} - D - (PLWHC \times D)} \quad (4.49)$$

where E is the excess liquid water [mm], Q_w is the liquid water available at the snow surface [mm], W_q is the liquid water held by the snow [mm], W_{qx} is the liquid water capacity [mm], D is the heat deficit [mm] and PLWHC the percent liquid water holding capacity [0.04]. If the water at the surface is only sufficient to overcome the heat deficit but not enough to fill the liquid water holding capacity, then the new liquid water is calculated by:

$$\boldsymbol{W_q = W_q + Q_w - D} \quad (4.50)$$

If there is not enough surface water to overcome the deficit, the deficit is reduced by the amount of available water Q_w and the amount of ice in the pack, W_i, is increased by the refreeze amount, Q_w, and the liquid water, W_q, remains the same.

4.5.3.8 Transmission of water through the snow cover

Excess liquid water must move through the snow before becoming outflow, and hence there is a time delay dependent on the snow condition. A portion of the excess water is withdrawn from the snowpack during a given time interval, governed by:

$$\boldsymbol{R_1 = \frac{1.0}{1.0 + 5e^{((-500E_{ls})/W_{is}{}^{1.3})}}} \quad (4.51)$$

where R_1 is the one hour withdrawal rate [mm], W_{is} is the mean water equivalent of the ice portion of the snow [inches] and, W_{is} is equal to W_i / 25.4 to convert mm to inches, E_{1s} is the average hourly lagged excess liquid water available for Δ [inches] and equal to E_l / 25 to convert mm to inches. E_1 is the average hourly lagged excess water available in the time-period [mm]. After lagging the excess water, the hourly outflow from the snow cover is:

$$\boldsymbol{O_{mr1} = (S_1 + E_1)R_1} \quad (4.52)$$

where O_{mr1} is the hourly snow cover outflow from melt or rain-on-snow [mm] and S_1 is the amount of lagged excess liquid water in storage at beginning of hour [mm], and the storage (S_2) at the end of the hour is equal to:

$$\boldsymbol{S_2 = S_1 + E_1 - O_{mr1}} \tag{4.53}$$

4.5.3.9 Heat transfer at the soil surface

Heat is transferred from the soil to the snow and vice versa depending on the temperature gradient in the lower layers of the soil. The amount of melt is small compared to that at the snow surface and calculated by:

$$\boldsymbol{M_g = DAYGM\left(\frac{\Delta t}{24}\right)} \tag{4.54}$$

where M_g is the amount of ground melt during each time-period [mm] and DAYGM is a model parameter representing ground melt [0.3 mm day^{-1}]. In addition to M_g, the outflow generated at the snow-soil interface, O_g, includes liquid water released due to a decrease in the amount of ice in the snowpack:

$$\boldsymbol{O_g = M_g + \left(\frac{M_g}{W_i}\right) W_q} \tag{4.55}$$

here O_g is the outflow due to ground melt for each time interval [mm], W_i is the mean water equivalent of the ice portion of the snow [mm], W_q is the ground melt parameter [0.3 mm day^{-1}] and W_i is the liquid water held by the snow [mm]. The total outflow from the snow cover is:

$$\boldsymbol{O_s = O_{mr} + O_g} \tag{4.56}$$

where O_s is the total snow cover outflow for each precipitation data interval [mm] and O_{mr} is the snow cover outflow from melt or rain-on-snow [mm].

4.5.4 Validation of the SNOW-17 model

Unlike the precipitation, temperature and evapotranspiration datasets developed within this chapter, the SNOW-17 model is not a stand-alone dataset that can be easily compared to

another for validation. As the SNOW-17 model will be integrated into the hydrological model, the snow simulation output is only available through running the full model. Although SWE datasets do exist on a daily scale (e.g. Canadian Meteorological Centre (CMC) Daily Snow Depth Analysis Data (Brown and Brasnet, 2010)), they are low resolution (25km grids) and hence difficult to compare to the snow layer of this hydrological model at a resolution of 15-arcseconds.

Therefore, calibration has not taken place on the SNOW-17 model to adjust model parameters, nor is the model validated directly. The model is added to give the snowfall/accumulation and melt a sense of reality based on its strong performance elsewhere within the literature.

4.6 Chapter conclusions

- River flow variation data across the YDB is extremely limited to a few specific points and hence to determine river flow variation across all rivers requires development of a distributed hydrological model.
- The sole input of water into a hydrological model of an independent catchment is through precipitation. Some of the water is lost to evapotranspiration before reaching the catchment outlet, largely influenced by temperature.
- Precipitation falls as rain or snow, requiring a snowfall/melt model, again largely influenced by temperature.
- Therefore, to develop a hydrological model requires precipitation, temperature and evapotranspiration datasets and a snowfall/melt model.
- Due to the fractal nature of precipitation in space and time, a high spatial and temporal resolution precipitation dataset is required, preferably at a global or continental scale.
- The APHRODITE dataset (Yatagai et al. 2009; 2012) was selected for precipitation due to its high quality, high resolution and long time-series.
- Daily rasters of the APHRODITE dataset were prepared from 1979 to 2007 to match MERRA data used for evapotranspiration as MERRA data only began in 1979 and 2007 is the end of the APHRODITE record.
- The YDB is characterised by wet summers and dry winters, with higher precipitation generally in the SE of China including the eastern half of the YDB.
- Precipitation varies annually but there is no evidence of a significant increasing or decreasing trend across the YDB between 1979 and 2007.

- For snowfall/melt modelling, the daily APHRODITE temperature dataset was also selected due to high quality, high resolution and long time-series.
- As minimum and maximum daily temperatures are required for evapotranspiration modelling, the MERRA data suite was selected due to its higher temporal resolution (albeit lower spatial resolution).
- The YDB is characterised by warm summers and cold winters, with higher temperatures on the central basin and eastern plains and cold temperatures on the western highlands.
- There is a significant rising trend of temperature over the period 1979 to 2007 which may be indicative of climate change and could influence future evapotranspiration, snowfall/melt and/or precipitation, impacting on hydropower development.
- Daily precipitation and temperature data is a compromise between model efficiency and data storage and accuracy, highlighted by the large diurnal temperature differences within the MERRA data suite.
- A unique daily distributed evapotranspiration dataset was created by the author based on the FAO method employing data within the MERRA data suite.
- As evapotranspiration is largely influenced by radiation (which influences temperature) the YDB is characterised by high evapotranspiration rates in summer and within the central basin and eastern plains and lower rates in the winter and over the western highlands.
- There is a significant rising trend in evapotranspiration rates over the period from 1979 to 2007, which could be indicative of future reducing river flow impacting on hydropower potential.
- Validation of the uniquely created evapotranspiration dataset resulting in a strong comparison with published monthly CRU data, with deviance assigned to the fact that crop factors were not incorporated to the created dataset.
- A sophisticated energy based conceptual index melt model was selected (SNOW-17) based largely on daily temperature due to its distributed nature and strong performance within the literature.
- Calibration and validation of the SNOW-17 model is not possible due to the lack of availability of snow based datasets across the YDB, but still selected due to its strong performance elsewhere within the literature.

5 Development of a hydrological model: Structure, calibration and outputs

5.1 Introduction

This chapter discusses the implementation of a hydrological model to determine surface runoff (river flow) variations across the YDB, incorporating terrain data (from chapter 3) and meteorological data (from chapter 4). Selection of an appropriate hydrological model depends on the input data available, resolution required and the ability to calibrate the model against observed river flow measurements. Calibration involves optimisation of model parameters with view to a good 'fit' between observed and model outputs, usually by aiming to optimise an objective function (efficiency criteria) by testing hundreds or thousands of parameter sets. Due to the size of the YDB, optimising parameter sets on sub-catchments is necessary that are then applied to the larger catchment. The aim of this modelling exercise is to generate rasters of daily flow, which are then combined to produce rasters representing percentile flow, where Q_n represents the flow which is equalled or exceeded n% of the time. These form inputs to the hydropower search algorithm.

5.2 An introduction to hydrological modelling

5.2.1 Classification of hydrological models

Pechlivanidis et al. (2011) reviewed catchment scale hydrological modelling types, calibration techniques and uncertainty analysis methods. They stated that due to resource constraints and limited range of measurement techniques, there are limitations to the availability of spatial-temporal data, and a need exists to extrapolate information from the available measurements in space and time.

In the same review, models were classified based on their structure. Metric models are primarily based on observations and aim to characterise the system response from available data, and are essentially empirical in nature. An example is Artificial Neural Networks (Lange, 1999; Jain et al., 2004; Dawson et al., 2006) that essentially 'learns' the relationship

between input (e.g. rainfall) and output (e.g. surface flow). Conceptual models specify the structure prior to modelling taking place and often the parameters do not have a direct physical interpretation (i.e. they are not independently measurable) and are estimated through calibration against observed data. In these models, a balance is often sought between the complexity of the model and the available data.

Physics based models represent the hydrological processes through the equations of motion based on continuum mechanics. In theory, they are based on measurable parameters and provide simulation of the runoff response without calibration. However, they are often based on small-scale field experiments that do not necessarily extrapolate to large catchments, and to reduce computational burden are often simplified adding to the questionability of the model (e.g. simplified St.Venant equations). Furthermore, catchments may have a high level of heterogeneity which can be difficult or even impossible to observe, particularly within a catchment the size of the Yangtze. Many models may be labelled as one of the above, but incorporate elements of two or more and hence are really hybrid models.

Models are further classified as lumped or distributed. Lumped models treat the whole catchment as a single unit, whereas distributed models make predictions distributed in space usually by discretising the catchment into elements (usually grid squares) and take advantage of spatial variability in input data and catchment characteristics. Semi-distributed models are a compromise of the two and the catchment is discretised as seen fit by the modeller, requiring less data and computational cost. Deterministic models produce a single set of results, whereas stochastic models use random variables to represent process uncertainty. Time based classifications include event based models, considering a single storm, or continuous simulation models considering a time-series of rainfall and sub-classified as sub-daily, daily, monthly etc. Singh and Singh (1995) classified models of catchments as small ($<100km^2$), medium ($100\text{-}1000km^2$) or large ($>1000km^2$). The Yangtze at ~ 2 million km^2 would possibly require its own classification.

5.2.2 Calibration of hydrological models

Calibration of models (also included in the Pechlivanidis et al. (2011) review) involves the selection of model parameters (process and physical) such that the hydrological response of the catchment is closely simulated. Physical parameters can be measured or estimated from data (e.g. surface slope, area) whereas process parameters cannot be normally measured (e.g.

depth of water storage capacity). Manual calibration involves adjusting parameters 'by hand' in a trial and error process, whereas automatic calibration is a computer-based method removing the subjective human judgement (Boyle et al., 2000). Results of each simulation are compared to an objective function (or goodness of fit), with the aim to maximise (or minimise as appropriate) the numerical value of the objective function.

The similarity between the model and observed data depends on the objective function selected, with different objective functions prioritising a particular hydrological response (i.e. prioritising goodness of fit with low flows, range of flows, mean flow etc). Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) is a popular choice but can underestimate variability and mean of flows (Pechlivanidis et al., 2010). Efficiency criteria that contain a summation of the square of the error term to avoid cancelling of errors of opposite sign (including NSE), emphasise larger errors and neglect smaller errors. Larger errors are usually associated with higher stream-flows. Therefore, calibration of a model (both manual and automatic) aimed at minimising NSE often leads to fitting higher portions of the hydrograph at the expense of baseflow.

Multi-objective analysis with a carefully selected number of objective functions depending on the modelling task, prevents a model being biased towards particular aspects of a hydrograph. A response surface is the surface described by the objective function in the parameter space, and optimisation algorithms search the response surface for parameter values that maximise the value of the objective function, constrained to pre-defined ranges of each parameter. Calibration is terminated when the objective function cannot be improved any further or when a maximum number of iterations have been completed.

Verification of the calibration results involves testing the parameters on a portion of data not used in the calibration. Often, the model response is better during calibration than verification, known as model divergence. When this divergence is unacceptable, the model structure and/or the calibration procedure requires modification.

5.3 Selection of an appropriate hydrological model

Kauffeldt et al. (2016) conducted a technical review of 24 large-scale hydrological models for implementation in operational flood forecasting schemes on a continental level. This review only considered those models with a clear large-scale focus, and hence are an ideal starting point for selection of a hydrological model suitable for modelling the Yangtze. The aim of

the paper was to select models suitable for the European Flood Awareness System (EFAS), but states that most of the requirements are considered universal for other continental or large-scale systems. Models were assessed on the following criteria: availability of the code; input data requirements; flexibility to grid structure; possibility of calibration; flexibility in resolution; facility to introduce discharge observation stations; pan-European model already set up (see table 5.1). A model suitable for the Yangtze requires these characteristics except for the latter (pan-European model already set-up).

Of the 23 models listed in table 5.1 (Land Dynamics Model excluded due to the code no longer maintained or available), 18 were rejected due to too coarse or too complex input data requirements, lack of calibration or lack of flexibility in resolution. Ideally the model would require input data requirements of daily temperature, precipitation and evapotranspiration datasets and operate on a grid resolution at least equal to the highest resolution of the HydroSHEDS dataset (i.e. 3-arcseconds or approximately 90m), although in this project it had already been decided to run the model at 15-arcseconds (approximately 450m).

Of the 5 models considered, LISFLOOD (De Roo et al., 2000) is primarily aimed at flood-plain modelling and is only available as an executable. The water balance WBMplus model (Wisser et al., 2010) was discounted due to its lower limit of resolution of 250m, which would prevent its usability for higher resolution grids (i.e. 3-arcseconds). SWAT (Soil and Water Assessment Tool) (Arnold et al., 1998; Neitsch et al., 2011) requires T_{min} and T_{max} as input which are not available in the high-resolution APHRODITE dataset. The mesoscale Hydrological Model (mHM) (Samaniego et al., 2010), although attractive, has only been applied at grid resolutions ranging from 1km to 100km, although it has been applied at basins up to 550,000km^2.

G2G, or Grid-to-Grid, developed through the Centre of Ecology and Hydrology (CEH) (Bell et al., 2007a,b) is a conceptual, distributed model able to simulate flows over large domains at high-resolution (it has been tested at 50m) and able to be integrated with GIS datasets. The code is only available as an executable file and developed for commercial use. However, the basic run off production and routing methodology was published (Bell et al., 2007). Although this would require formulation of new code, the characteristics of G2G made it a suitable choice for modelling of the Yangtze basin.

Duncan (2014), due to the features of G2G, used these published equations to develop a model to produce daily flows in an assessment of Scotland's hydropower, however, this code,

developed in Python and C++, is not available. Hence using the published equations, a new model was developed in R and C++ specifically for this project, integrating the datasets developed in chapters 3 and 4.

Model	Code Available	Input data requirements	Flexible to grid	Calibration Possible	Flexible in resolution	Introduce discharge station observations?
CLM	Open	<6h, ~2/0.1° S, R, SR, LW, SP, Q, T, W, aerosol, CO2	Yes	No	5km to 10km	Yes
E-HYPE	Open	D, 215km^2, P, T	Yes	Yes	Yes	Yes
G2G	Exe only	<15min, <1km, P, T, PET	Yes	Yes	Yes	Yes
GWAVA	Exe only	D/M, 0.5°, P, T, PET	Yes	Yes	>10km	Yes
H08	Open	6h, 1°/0.5, R, S, T, Q, W, SP, SW, LW	Yes	No	Yes	No
H-TESSEL	Free licence	1h, >0.25°, R, S, T, Q, SP, W, SW, LW	Yes	No	Yes	No
JULES	Open	1h, 1-50km, R, S, T, W, Q, LW, SW, SP	Yes	No	Yes	No
LISFLOOD	Exe only	6h/D, 5km, P, PET, T	Yes	Yes	Yes	Yes
LPJml	Not available	D, 0.5°, P, T, LWnet, SW	Yes	No	0.5°	No
Mac-PDM	Yes	D/M, 0.5°/20km, P, T, W, Q, LWnet, SW	Yes	No	Yes	No
MATSIRO	On request	1h, 0.5°/1°, R, S, T, W, Q, LW, SW, SP	Yes	No	>50km	No
mHM	Open	1h/D, 1-100km^2, P, T, PET	Yes	Yes	Yes	Yes
MPI-HM	Scientific licence	D, 0.5°, P, T	Yes	No	0.5°	No
NOAH-MP	Open	3h, 0.5°, P, T, SW, LW, Q, SP, W	Yes	No	Yes	No
ORCHIDEE	Open	15min/3h, 0.5°/1°, P, T, W, SR, Q, CO_2	Yes	No	Yes	No
PCR-GLOBWB	Open	D, 0.5/5', S, P, T, PET, W, GR, Q	Yes	Yes	>30°	Yes
SWAT	Open	D, 1h-3000km^2, P, Tmin, Tmax	Yes	Yes	Yes	Yes
SWIM	Upon request	D, 1000 – 500,000km^2, P, T, SR, Q	Yes	Yes	Yes	No
TOPLATS	Open	1-3hr, 10m-1km, P, T, LW, SW, SP, Q, W	Yes	Yes	10m – 1km	Yes
VIC	Open	1h-D, 6-222km, P, Tmin, Tmax, W	Yes	Yes	>6km	No
WASMOD-M	Not available	D/M, 0.25/0.5°, P, T, PET	Yes	No	0.25°	No
WaterGAP	Not available	D/M, 0.5°/5', p,T, cloud/GR, LW	Yes	No	10km	Yes
WBMplus	Open	D, 0.5°, P, T	Yes	Yes	250m to 2°	Yes

Table 5.1: Characteristics of hydrological models suitable for large-scale catchment modelling (for key see Kauffeldt et al., 2015). Green = suitable for hydrological modelling of the Yangtze

5.4 A description of the G2G hydrological model

5.4.1 An introduction to the G2G hydrological model

The G2G model was developed for regional assessments of the impact of climate change on river flow systems and to assess the land-surface component of a regional climate model. G2G is a spatially-distributed hydrological runoff production and routing scheme for use with gridded atmospheric data, configured spatially using river networks and terrain information derived from a DEM. A modelling framework is provided to translate estimates of current or future rainfall into area-wide estimates of river flow at a daily/sub-daily timestep. Importantly for this project, the model can provide river flow estimates for areas that are ungauged. A representation of the G2G model is given in figure 5.1.

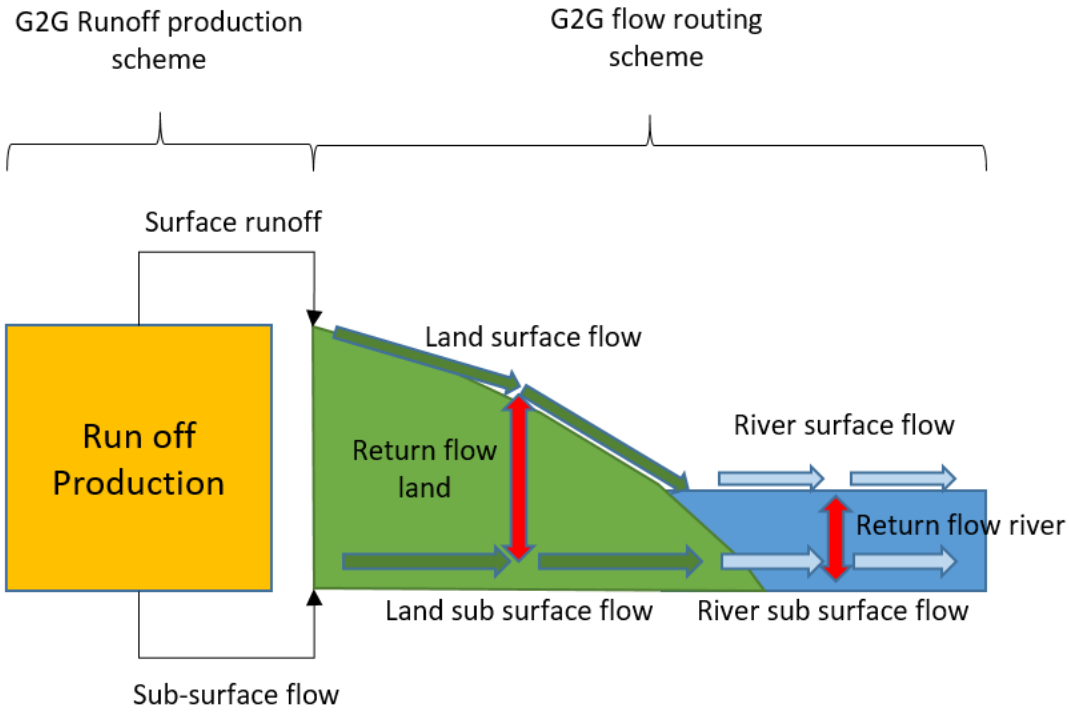


Figure 5.1: Overview of the G2G model (Bell et al., 2007)

5.4.2 The G2G runoff production scheme and probability-distributed soil moisture

G2G employs a runoff production scheme, based on the runoff production component of the CEH Grid Model (Bell and Moore, 1998), to provide grid-based estimates of surface and sub-

surface runoff to the G2G routing model. For each grid square, the maximum storage capacity is linked to the average topographic gradient (derived from a DEM) by:

$$\mathbf{S_{max}} = \left(\mathbf{1} - \frac{\bar{\mathbf{g}}}{\mathbf{g_{max}}}\right)\mathbf{c_{max}} \qquad (5.1)$$

where S_{max} is the maximum storage capacity [mm], $\bar{g}$ is the average topographic gradient (derived from the DEM), g_{max} is the regional upper limit of topographic gradient and c_{max} is the regional upper limit of storage capacity (a calibrated parameter).

Each grid square can be imagined as a 'leaky bucket'. If a grid square is saturated due to previous rainfall events or incoming surface runoff, then further rainfall spills over contributing to surface runoff (termed the fast catchment response). Drainage from the base of the grid square enters baseflow (the slow catchment response) and depends on the volume of water stored in the grid square. Water is also lost due to evaporation from the top of the column and a water balance is maintained for each grid square. A diagram representation is given in figure 5.2.

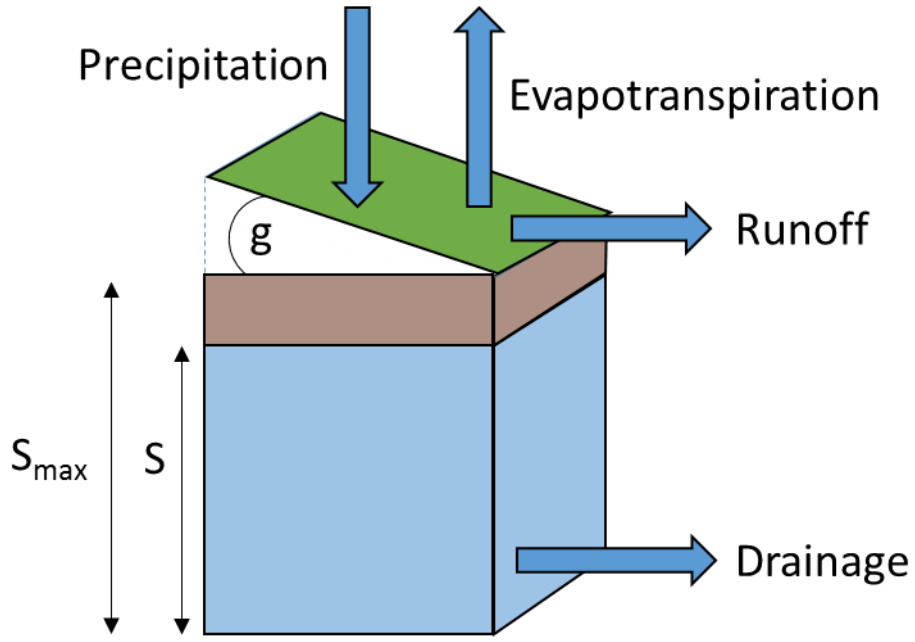


Figure 5.2: Representation of the grid box water balance (Bell et al., 2007a); S = Current grid box storage; Smax = Maximum grid box storage; g = slope

Evaporation loss from the soil column occurs at the rate E_a, related to the reference crop evapotranspiration rate (ET_o) per hour, E, by:

$$\mathbf{E_a = E\left\{1 - \left(\frac{S_{max} - S}{S_{max}}\right)\right\}} \quad (5.2)$$

where E_a is evaporation loss from the soil column in the timestep [mm], S_{max} is the maximum storage capacity [mm], E is the reference crop evapotranspiration (ET_o) within the timestep [mm] and S is the current depth of water in store [mm].

Drainage from the base of the grid occurs at the rate:

$$\mathbf{d} = \begin{cases} \mathbf{k_d S^{\beta}}, & \mathbf{S > 0} \\ \mathbf{0}, & \mathbf{S \leq 0} \end{cases} \quad (5.3)$$

where d is drainage from the base of the grid within the timesetep [mm], k_d is the storage rate constant [mm^{-2}] and β is a parameter but here set to 3 [dimensionless]. The potential water storage is given by:

$$\mathbf{S = max(0, S + p\Delta t - E_a \Delta t - d\Delta t)} \quad (5.4)$$

where p is precipitation within the timestep [mm] and Δt is the timestep [one hour]. The water storage S is reset to S_{max} if direct runoff is generated and the direct runoff contributing to the fast catchment response, q [mm] is:

$$\mathbf{q = max(0, S - S_{max})} \quad (5.5)$$

Soil water storage heterogeneity within a grid square is introduced by employing the probability-distributed soil moisture (PDM) formulation developed by Moore (1985). It assumes that a proportion of the grid square is saturated and generating runoff even when rainfall amounts are small, otherwise the whole grid square would have to be saturated before runoff was generated. This critical moisture value (Ct) is found by:

$$\mathbf{Ct = c_{max} - \left(c_{max} * \left(1 - \frac{S * (b + 1)}{c_{max}} \right)^{\frac{1}{b+1}} \right)} \quad (5.6)$$

where c_{max} is the regional maximum storage capacity [mm] and b is a parameter determined from grid square slope values where:

$$\mathbf{b} = \frac{\bar{\mathbf{g}}}{\mathbf{g_{max}} - \bar{\mathbf{g}}} \tag{5.7}$$

where $\bar{g}$ is the grid square slope and g_{max} is the regional maximum slope.

5.4.3 The G2G flow routing scheme

Flow is routed in the G2G model based upon a discrete approximation to the 1-D kinematic wave equation with lateral inflow, an approximation to the St.Venant equations for surface flows. Kinematic routing is applied to surface and sub-surface flows and allows for different formulations over land and river pathways. Flow transfers between the surface and sub-surface is allowed for with a *return flow* term, a spatially continuous method of combining the fast and slow catchment responses. The model equations are:

$$\frac{\partial \mathbf{q_l}}{\partial \mathbf{t}} + \mathbf{c_l}\frac{\partial \mathbf{q_l}}{\partial \mathbf{x}} = \mathbf{c_l}(\mathbf{u_l} + \mathbf{R_l}) \tag{5.8}$$

$$\frac{\partial \mathbf{q_{lb}}}{\partial \mathbf{t}} + \mathbf{c_{lb}}\frac{\partial \mathbf{q_{lb}}}{\partial \mathbf{x}} = \mathbf{c_l b}(\mathbf{u_{lb}} - \mathbf{R_l})$$

$$\frac{\partial \mathbf{q_r}}{\partial \mathbf{t}} + \mathbf{c_r}\frac{\partial \mathbf{q_r}}{\partial \mathbf{x}} = \mathbf{c_r}(\mathbf{u_r} + \mathbf{R_r})$$

$$\frac{\partial \mathbf{q_{rb}}}{\partial \mathbf{t}} + \mathbf{c_{rb}}\frac{\partial \mathbf{q_{rb}}}{\partial \mathbf{x}} = \mathbf{c_{rb}}(\mathbf{u_{rb}} - \mathbf{R_{rb}})$$

where R is return flow, q is channel flow, u is lateral flow, c is the kinematic wave speed, l and r denote land and river respectively and b denotes sub surface flows. Dividing t and x into discrete intervals and invoking forward difference approximations to the derivatives gives the discrete formulation:

$$\mathbf{q_k^n} = (\mathbf{1} - \boldsymbol{\theta})\mathbf{q_{k-1}^n} + \boldsymbol{\theta}(\mathbf{q_{k-1}^{n-1}} + \mathbf{u_k^n} + \mathbf{R_k^n}) \tag{5.9}$$

where θ is the dimensionless wave speed equal to $c\frac{\Delta t}{\Delta x}$ and $0 < \theta < 1$, and n and k denote positions in discrete time and space. This is a simple explicit numerical formulation which introduces diffusion representing the propagation of actual rivers, and used to develop routing for land and river flows. For application to two dimensions, the q_{k-1}^{n-1} term represents inflow from the preceding grid cell in space and is given by the sum of the inflows from adjacent grid-cells. In practice the routing is implemented in terms of the equivalent depth of water in store over the grid square, S_k^n, where:

$$\mathbf{q_k^n = \kappa S_k^n \text{ where } \kappa = \frac{c}{\Delta x}} \qquad (5.10)$$

where S_k^n is runoff/sub-surface flow over/under the grid square [mm], q_k^n is channel flow of the cell at the current timestep [mm s^{-1}], κ is a rate constant [s^{-1}], c is the wave celerity [ms^{-1}] and Δx is the size of the grid square [m]. The inflows to the grid square u_r, u_l, u_{rb} and u_{lb} are represented by the surface and sub-surface terms q and d in the runoff production scheme.

5.4.4 G2G model parameters requiring calibration

From sections 5.4.2 and 5.4.3, the G2G model has 8 parameters requiring calibration, with values given by Bell et al. (2007a) shown in table 5.2.

Parameter	Symbol	Units	Value in Bell et al. (2007a)
Surface wave speed land	c_l	ms^{-1}	0.4
Surface wave speed river	c_r	ms^{-1}	0.5
Sub-surface wave speed land	c_{lb}	ms^{-1}	0.05
Sub-surface wave speed river	c_{rb}	ms^{-1}	0.05
Return flow land	R_l	-	0.005
Return flow river	R_r	-	0.005
Maximum grid cell storage	c_{max}	mm	140
Drainage storage rate constant	k_d	-	0.00005

Table 5.2: Parameters requiring calibration in the G2G model

5.5 Structure of the model code

The model code for this project was mostly developed in the R programming language with some of the matrix calculations written in C++. R is an interpreted language with many supporting libraries, making code writing accessible, but suffers in terms of speed compared to lower level languages such as C++. However, the extra development time coding in C++ may have offset any speed benefits.

The code was written as a series of functions that are called when appropriate during the running of the code, illustrated in figure 5.3. Development and testing was carried out on a local Intel Core-i7 PC with 32GB of memory running Microsoft R Open 3.2.3 64 bit on a Windows 10 platform. This was then modified to Linux code to run on the Edinburgh (University) Compute and Data Facility, known as Eddie, a high-performance computing

environment enabling access to multiple CPUs and parallel programming. Eddie Linux Compute Cluster Mark 3 is the third iteration of the University's compute cluster consisting of 4000 Intel Xeon cores with up to 2TB of memory per compute node. Most nodes are Intel Xeon Processor E5-2630 v3 (2.4GHz), and programs ran via the Open Grid Scheduler/Grid Engine (Sun Microsystems) on Scientific Linux 7.

5.6 A description of the hydrological modelling code

5.6.1 Installing packages and loading libraries

When running the code for the first time, several R packages need to be installed which are available as ".tar" compressed files available via the Comprehensive R Archive Network (CRAN) (https://cran.r-project.org/). Different versions are available for Linux and Windows. Once installed on a machine or server, they do not require installing again. *RGDAL* (Bindings for the Geospatial Data Abstraction Library) (Bivand et al., 2017a) and *RGEOS* (Interface to Geometry Engine – Open Source) (Bivand et al., 2015) are both required but need administrator installation on the Eddie system.

The R packages installed are then loaded as libraries in the *Load Libraries* function. Requisite libraries include:

- *maptools* (Bivand et al., 2017b) - Tools for reading and handling spatial objects: a set of tools for manipulating and reading geographic data, in particular 'ESRI Shapefiles'
- *raster* (Hijmans et al., 2016a) - Geographic data analysis and modelling: reading, writing, manipulating, analysing and modelling of gridded spatial data
- *Rcpp* (Eddelbuettel et al., 2017) - Seamless R and C++ integration: provides R functions as well as C++ classes which offer a seamless integration of R and C++. Many R data types and objects can be mapped back and forth to C++ equivalents which facilitates both writing of new code as well as easier integration of third-party libraries
- *Parallel* (available within base R after release 2.14.0) – incorporates (slightly revised) copies of previous packages *MULTICORE* and *SNOW* (Simple Network of Workstations) with support for simple parallel computing in R

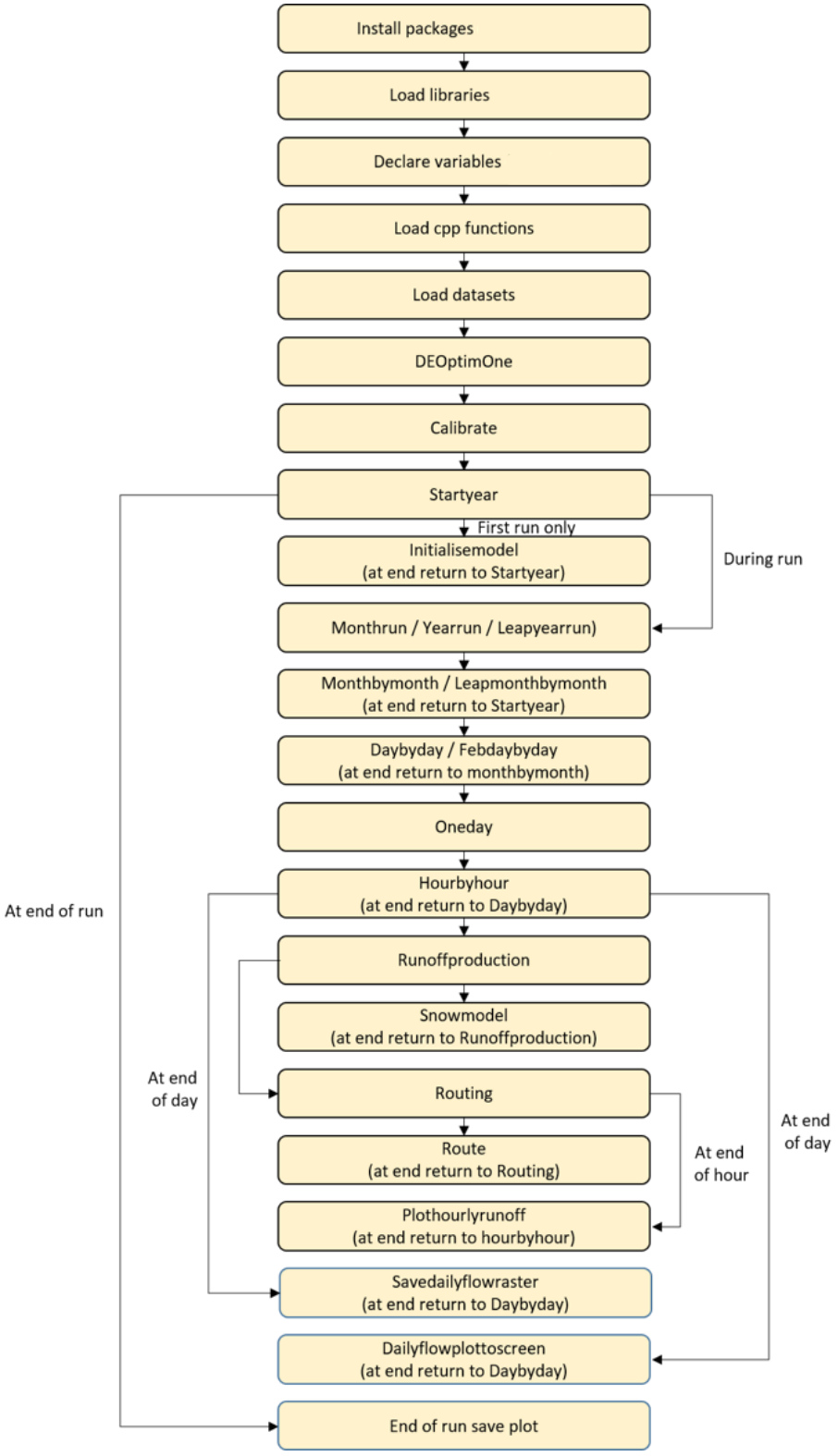


Figure 5.3: Overview of modelling code functions

- *Rcpp Armadillo* (Eddelbuettel et al., 2017) - 'Rcpp' integration for the 'Armadillo' templated linear algebra library - a templated C++ linear algebra library (by Conrad Sanderson) that aims for a good balance between speed and ease of use. Integer, floating point and complex numbers are supported, as well as a subset of trigonometric and statistics functions. Various matrix decompositions are provided through optional integration with *LAPACK* and *ATLAS* libraries.
- *DEoptim* (Ardia et al., 2016) - Global optimization by differential evolution: implements the differential evolution algorithm for global optimization of a real-valued function of a real-valued parameter vector. *DEOptim* is the method selected to calibrate the model and discussed in 5.10.

5.6.2 Declaring variables

Calibration of the model requires hundreds/thousands of model runs to test different combinations of the parameter set and optimise the objective function. Parallel computing enables each core of a machine to run the model with a different set of parameters, effectively increasing the efficiency and speed of calibration. The *Declare variables* function is required to load all the used variables within the code (approximately 550) which are then passed to the *DEoptimone* function and subsequently each new core utilised.

5.6.3 *Load C++* functions

Load cpp functions creates Rcpp and Rcpp Armadillo C++ code to manipulate matrices faster than what raw R could achieve. R raster files are a memory efficient way of storing gridded data with the advantage that they also contain spatial extent information, but are slow to manipulate in calculations. Therefore, all raster files are converted to matrices within the code which are faster in calculations, but comes at the expense of greater memory requirements and matrices do not contain spatial information.

Three functions are created entitled *EastC, WestC* and *MoveC*, designed to move specific cells within matrices in the directions specified by the FDR. These are implemented in the *routing* part of the model. The Rcpp function *EastC* moves those cells that are designated to move east (or southeast or northeast) by shifting a matrix of those cells one step to the right (i.e. east) by creating a copy but where the column number, j, of the new matrix is equal to j-1 of the original matrix and the first column set to a one column matrix of zeros (see figure 5.4).

WestC is similar but shifts those cells that move west (or southwest or northwest) one step to the left and the column of zeros is added at the last column. Those cells that fall outside the extent of the original matrix extent are removed completely (this usually only happens when flow moves out of the catchment extent).

Original matrix

1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	1	0	0
0	1	0	0	0	0	1	0	1	0
0	0	1	0	0	1	0	1	1	0
0	0	0	1	0	0	1	0	0	1
0	1	0	0	1	1	0	0	1	1
0	0	0	1	0	0	0	1	0	1
1	0	1	0	0	0	1	0	1	0
0	1	0	1	0	1	0	0	0	0

EastC function

Final Matrix – all the cells have moved east

1	2	3	4	5	6	7	8	9	10
0	1	0	0	0	0	0	0	1	0
0	0	1	0	0	0	0	1	0	1
0	0	0	1	0	0	1	0	1	1
0	0	0	0	1	0	0	1	0	0
0	0	1	0	0	1	1	0	0	1
0	0	0	0	1	0	0	0	1	0
0	1	0	1	0	0	0	1	0	1
0	0	1	0	1	0	1	0	0	0

Figure 5.4: Example application of the Rcpp EastC function, moving all cells one step to the right (east) where 1 in the original matrix denotes those cells that move east

The *MoveC* function is Rcpp Armadillo code designed to move those cells that move south (or southeast or southwest) one row down and inserts a row of zeros in the first row, and similarly move those cells that move north (or northeast or northwest) one row up and inserts a row of zeros in the last row. *MoveC* then combines (adds) all the moved matrices (east, west, south, north, southeast, southwest, northeast and northwest) to compute where the runoff (or baseflow) has moved to.

5.6.4 Loading datasets

The *Load datasets* function reads in the catchment terrain datasets as determined earlier, including the 15-arcseconds FDR, DEM, slope and river network rasters, and shapefiles of the catchment and the co-ordinates of a calibration measurement station (i.e. a single point exported from ArcGIS). The reference measurement station does not necessarily have to be at the catchment outlet. To decrease the 'warm-up' time of the catchment, starting conditions are initialised by reading in rasters representing the water storage of each cell, initial snow conditions, initial baseflow and initial runoff. *SStart* (figure 5.5), *Snowstart* (figure 5.6) and *Baseflowstart* (figure 5.7) are gridded October mean estimates of storage, snow and baseflow respectively across the YDB extracted from the MERRA data suite variables FRSAT (Fractional saturated area), SNOMAS (Snow mass) and BASEFLOW (Baseflow) located within the 2D land surface diagnostics at the native resolution of 0.5° x 0.667°, and up-scaled to the 15-arcseconds resolution. A simple estimate of initial land storage conditions was calculated by:

$$\textbf{Initial storage } [\textbf{mm}] = \left(\mathbf{1} - \frac{\bar{\mathbf{g}}}{\mathbf{g_{max}}}\right)\mathbf{c_{max}} * \textbf{FRSAT} \qquad (5.11)$$

where $\bar{g}$ is the average topographic gradient (i.e. slope), g_{max} is the regional upper limit of topographic gradient (maximum slope), c_{max} is the regional upper limit of storage capacity, set to 100mm, and FRSAT is the MERRA fractional saturated area [a fraction]. Note that the units of kgm^{-2} are interchangeable with mm of water i.e. a snow mass of $10kgm^{-2}$ would be equivalent to a depth of 10mm of water over the same area. Similarly, baseflow units of $kgm^{-2}s^{-1}$ can be represented as mms^{-1}, hence the model accounts for all water inputs and outputs in mm. As the initial resolution of the MERRA data is course, the *Baseflowstart* raster does not represent well the baseflow found in river cells.

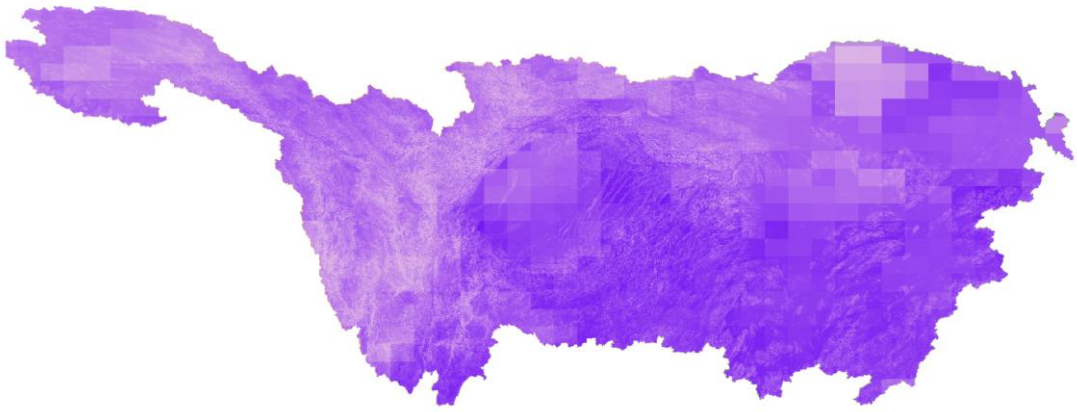


Figure 5.5: MERRA estimated initial storage conditions ranging from near to 0mm (very light purple) to 60mm (dark purple). Steep slopes contain less initial storage water.

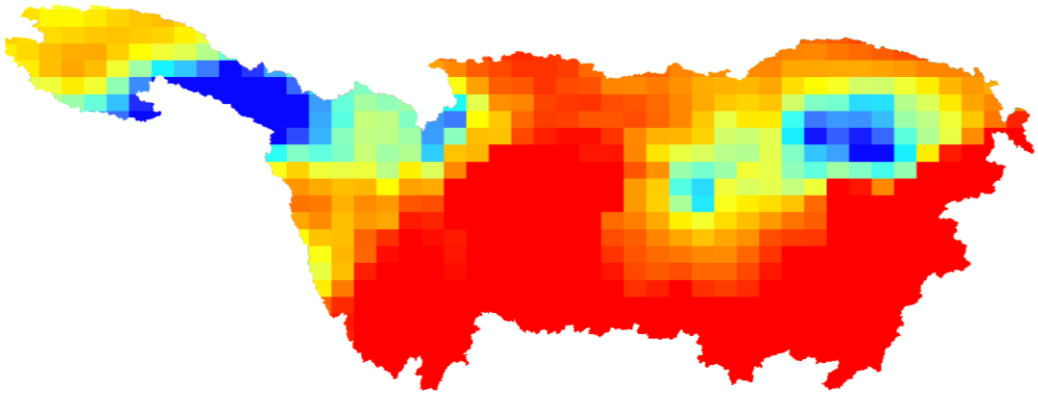


Figure 5.6: MERRA estimated initial snow mass conditions ranging from near to 0 (red) to 47 kgm^{-2} (dark blue). Highland areas generally contain more initial snow mass.

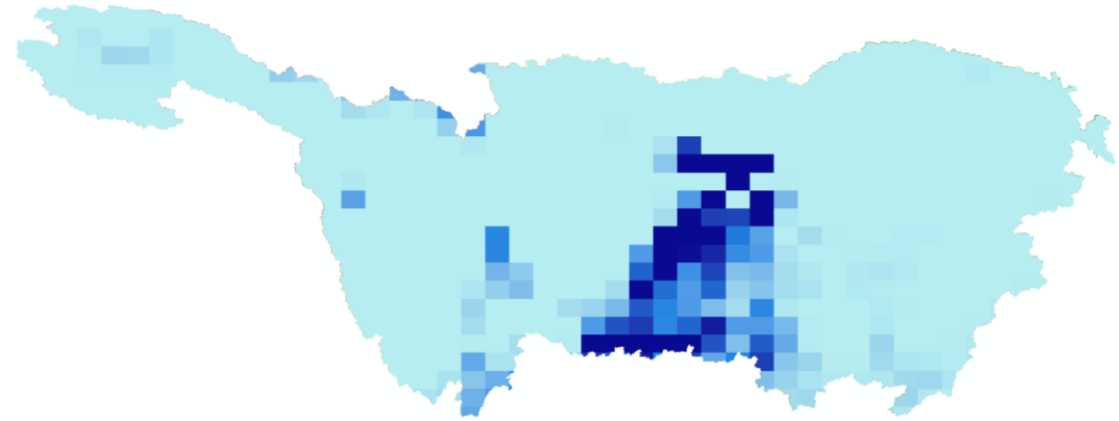


*Figure 5.7: MERRA estimated initial baseflow conditions ranging from near to 0 (light blue) to 1.9*10^{-5} kgm^{-2}s^{-1} (dark blue)*

A weighted runoff raster was developed in ArcGIS by using the flow accumulation tool on the YDB weighted by the GRDC Global Composite Fields data in mms^{-1} (Fekete et al., 2002) (discussed in 3.6). This gives the mean annual surface runoff across the YDB. Average flow at the Yangtze mouth calculated using this method is approximately 30,000m^{3}s^{-1}, which is similar to that reported by other sources.

The raster of river cells was derived from the weighted runoff dataset with a minimum annual average flow of 5ls^{-1}, used to set the wave celerity (as there are different parameter values for river and non-river) and to modify the initial storage conditions of river cells to fully saturated.

Co-ordinates of the shapefile of the point at the catchment outlet or measurement reference station are extracted and converted to a raster of one cell at 15-arcseconds resolution and 'snapped' (aligned) to the FDR raster. A raster representing catchment-wide atmospheric pressure (see figure 5.8) required as part of the snowmelt model (see section 4.5.3) is calculated by the barometric formula:

$$\mathbf{P_a} = \frac{\mathbf{P_0}\ \mathbf{exp}\left(-\frac{\mathbf{gMh}}{\mathbf{R_0 T_0}}\right)}{\mathbf{100}} \qquad (5.12)$$

where P_a is the atmospheric pressure at DEM altitude [mb], P_0 is the atmospheric pressure at sea-level [101325Pascals], g is the acceleration due to gravity [9.80665ms^{-2}], M is the molar mass of dry air [0.0289644kg mol^{-1}], h is the DEM elevation [m], R_0 is the universal gas constant [8.31447 J mol^{-1}K^{-1}] and T_0 is the sea level standard temperature [288.15K].

Figure 5.8: Atmospheric pressure [mb] across the YDB calculated by the barometric formula

Finally, the area of the catchment is calculated by finding the area of each cell within the catchment (using the Raster package *area* tool). The height of each cell is fixed whatever the latitude, but the width decreases as the latitude moves north away from the equator and hence the area decreases moving north along the catchment (see figure 5.9):

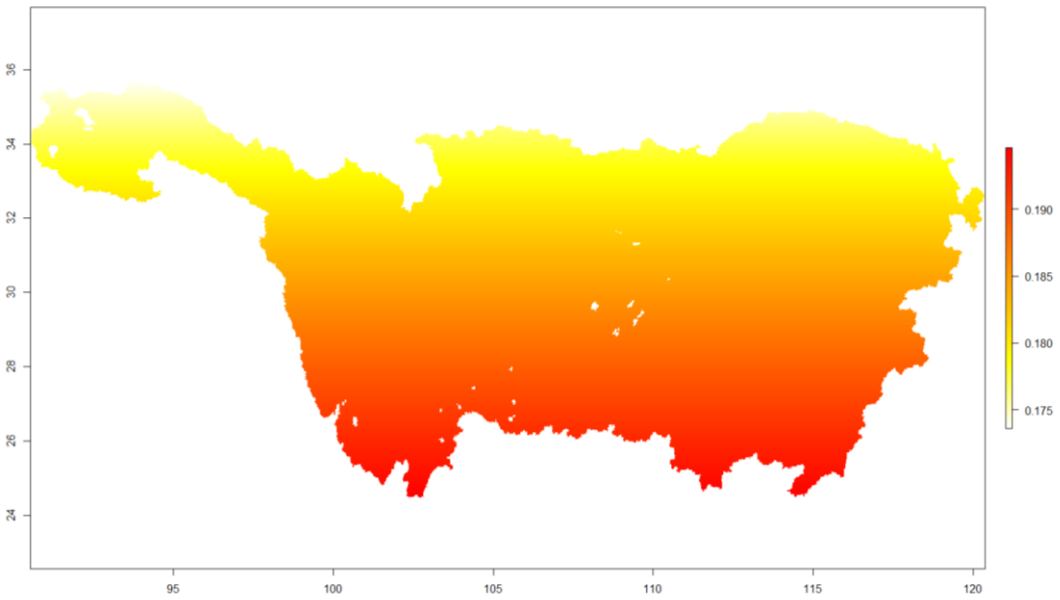


Figure 5.9: Area of each cell [km^2] across the YDB calculated using the raster area tool, ranging from 0.1946km^2 in the south to 0.1736km^2 in the north

The area of individual cells is summed to give a total, calculated as 1,902,401km^2 for the YDB. As the cell height is fixed (0.4633km for a 15-arcsec cell), the cell width is calculated

by dividing the area by the cell height and assumes the cell is fixed width across the cell. The diagonal length is then calculated by:

$$\textbf{Diagonal length} = \sqrt{\textbf{height}^2 + \textbf{width}^2} \qquad (5.13)$$

5.6.5 The DEOptim function

*DEOptimeOn*e is a re-write of the original *DEOptim* script (see section 5.10) to enable flexibility over running the model in parallel. This was added to enable selection of the number of parallel cores to perform the calibration and an additional statement to enable parallel processing on a Linux machine, where type = "FORK":

```
cl <- parallel::makeCluster(parallel::detectCores(),type="FORK")
```

When the *DEOptimOne* function is called, all the variable names declared in *Declare Variables* (section 5.6.2) and all the R packages necessary (section 5.6.1) are passed to any cores taking part in the calibration. This subsequently calls the *Calibrate* function (section 5.6.6) selecting the 8 calibration parameters (see section 5.9) between a *lower* and *upper* value (*see section 5.10* for further details about calibration).

The *DEOptimOne* function's purpose is to optimise an objective function, aiming to select parameters so the objective function is as near to 1 as possible i.e. aiming for a perfect match between modelled and observed river flows at the calibration station. The objective function choice is flexible but in this project multi-objective efficiency criteria are selected (see section 5.8).

5.6.6 The calibrate function and setting the start date

The *Calibrate* function starts the hydrological model by setting the desired year and month to start modelling and initialises the model by running the *Initialisemodel* function. Usually the model is 'warmed up' by running the model for a few months before the desired start month e.g. if desiring to start the model in January 1980, then the model will first run the months of October, November and December in 1979. After the warm-up is completed, the year of the desired run is selected, the month set to January and the *Yearrun/Leapyearrun* function is initiated (a 365 or 366-day year). Subsequent years are modelled by changing the year and running *Yearrun* or *Leapyearrun* again.

5.6.7 Initialising the model

The model is initialised by setting river cells in the wave celerity raster to parameter c_r and non-river cells to parameter c_l. Similarly, a baseflow celerity raster is set to c_{rb} and c_{lb} and a return flow raster set to parameters r_r and r_l if river or non-river respectively. The wave-speed of the model is then determined as:

$$\mathbf{Wavespeed} = \mathbf{max}(\mathbf{c_r}, \mathbf{c_l}) \qquad (5.14)$$

where c_r is the surface wave speed in rivers [ms^{-1}] and c_l is the surface wave speed over non-river land [ms^{-1}]. This is used to determine the timestep in minutes, which must be small enough to ensure flow only moves into the next cell as determined by the flow direction, and not subsequent cells. The timestep of the model is then determined by:

$$\mathbf{Timestep\ [minutes]} = \mathbf{Floor}\left[\frac{\mathbf{minimum\ cell\ width\ [m]}}{\mathbf{60} * \mathbf{wavespeed\ [ms^{-1}]}}\right] \qquad (5.15)$$

Floor returns the largest integer not greater than the calculation in the brackets in equation 5.15. To enable the timestep to be replicated a number of times per hour, 60/timestep must be a whole number, and hence the timestep is determined by an 'if' statement as follows:

If statement [minutes]	Result [minutes]
If Timestep = 1 or 2 or 3 or 4	Timestep = Timestep
If Timestep = 6 or 7 or 8 or 9	Timestep = 6
If Timestep = 10 or 11	Timestep = 10
If Timestep = 12 or 13 or 14	Timestep = 12
If Timestep >= 15 and < 20	Timestep = 15
If Timestep >= 21 and <30	Timestep = 20
If Timestep >= 30 and <=59	Timestep = 30

Table 5.3: 'If' statements for calculation of the model timestep [minutes]

The number of times the model is replicated per hour is determined by:

$$\mathbf{Replication} = \frac{\mathbf{60}}{\mathbf{Timestep}} \qquad (5.16)$$

River cells are set to saturated at the model start (i.e. S_{max}). To ensure flow moves in the correct direction, rasters are created representing those cells that move in one of the eight cardinal directions, and used to develop flow directional Rate Constant [s^{-1}] rasters:

$$\textbf{Rate Constant (E or W)} = \frac{\textbf{Wave Celerity } [\textbf{ms}^{-1}]}{\textbf{Cell width } [\textbf{m}]} \quad (5.17)$$

$$\textbf{Rate Constant (N or S)} = \frac{\textbf{Wave Celerity } [\textbf{ms}^{-1}]}{\textbf{Cell height } [\textbf{m}]} \quad (5.18)$$

$$\textbf{Rate Constant (NE, NW, SE, SW)} = \frac{\textbf{Wave Celerity } [\textbf{ms}^{-1}]}{\textbf{Cell diagnaol } [\textbf{m}]} \quad (5.19)$$

These are recombined into a *Rate Constant* raster for the whole catchment. The amount of water remaining in a cell after each timestep is calculated using:

$$\textbf{Remaining surface rate constant} = \textbf{1} - \textbf{Rate constant} \quad (5.20)$$

Similar rasters are created for baseflow, and the return flow rate is calculated by:

$$\textbf{Return flow rate constant} = \textbf{Return flow x Baseflow rate constant} \quad (5.21)$$

At this stage, all the created rasters are converted to *matrix* format as discussed in section 5.6.3. Matrices are converted back to raster format for plotting and saving within other functions.

5.6.8 Starting the model

Monthbymonth and *Leapmonthbymonth* functions set the starting month as January and then calls the *daybyday* function representing 30 days, and repeats for subsequent months. For February, the *febdaybyday* function is called (i.e. as February only has 28 days). For months with 31 days, an extra day is modelled (day 31). In the *leapmonthbymonth* function an extra day is added to February.

5.6.9 Setting the meteorological environment for each day

In the *daybyday/febdaybyday* functions, the day is set to "01" and the *oneday* function is called at the start of the day, and repeated for each day of the month. A daily evapotranspiration raster is loaded into the model based on the current year, month and day and cropped to the catchment. As the loaded evapotranspiration file is at the APHRODITE resolution of 0.25°x0.25°, the raster is disaggregated by 60 to convert to a raster of 15-arcseconds resolution. Disaggregation splits the original cells into smaller cells but keeps the underlying

0.25°x0.25° data for each cell. The raster is then divided by 24 to establish an hourly evapotranspiration rate. Precipitation data is also loaded and processed in the same manner.

Temperature data is also loaded and cropped to the catchment. As the original temperature data is at 0.25°x0.25°, the temperature data is increased in resolution to 15-arcseconds and adjusted to the altitude of the 15-arcseconds DEM. The method to achieve this is as follows:

1. The 15-arcsecond DEM (see figure 5.10) is aggregated by 60, with the resulting data taking the mean of the original 60 underlying cells of the 15-arcseconds DEM (aggregation is the reverse of disaggregation i.e. many cells are combined to one cell), producing a DEM of 0.25°x0.25° resolution.
2. The 0.25°x0.25° DEM is then disaggregated by 60 to produce a 15-arcsecond DEM but with the underlying data of the coarser DEM (see figure 5.11).
3. The difference in elevation is found between the original 15-arcseconds DEM and the 15-arcseconds DEM found in step 2.
4. The temperature difference is found by multiplying the DEM difference found in step 3 by the adiabatic lapse rate (6.49°C/1000m) (see figure 5.12).
5. The final temperature raster (see figure 5.14) is the original 0.25°x0.25° temperature raster (see figure 5.13) disaggregated by 60 (i.e. a 15-arcsecond temperature raster) minus the temperature difference raster in step 4.

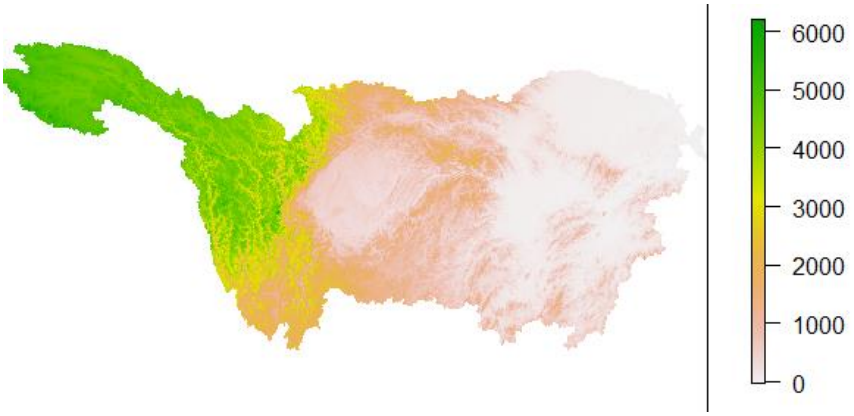


Figure 5.10: Original DEM raster at 15-arcseconds [m]

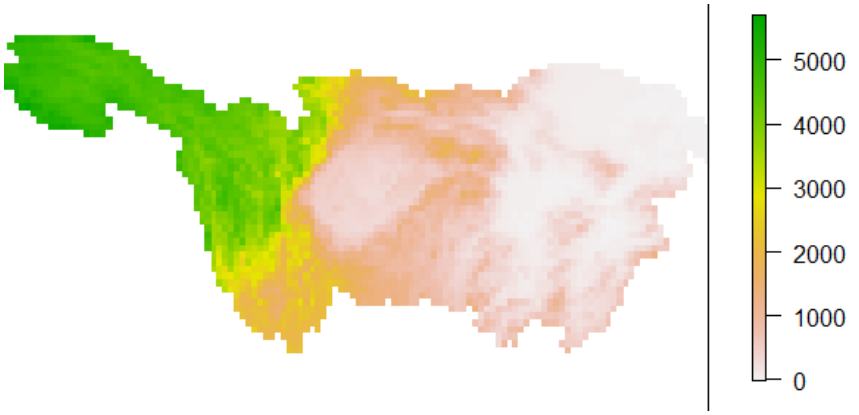


Figure 5.11: DEM raster at 15-arcseconds following aggregation (by 60) and then disaggregation (by 60) [m]

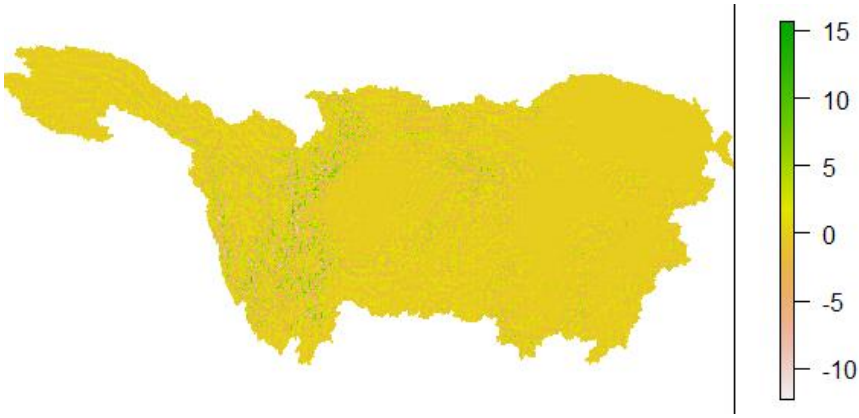


Figure 5.12: Difference in temperature between original temperature raster at 0.25°x0.25° resolution and 15-arcsecond temperature raster adjusted by DEM difference [°C]

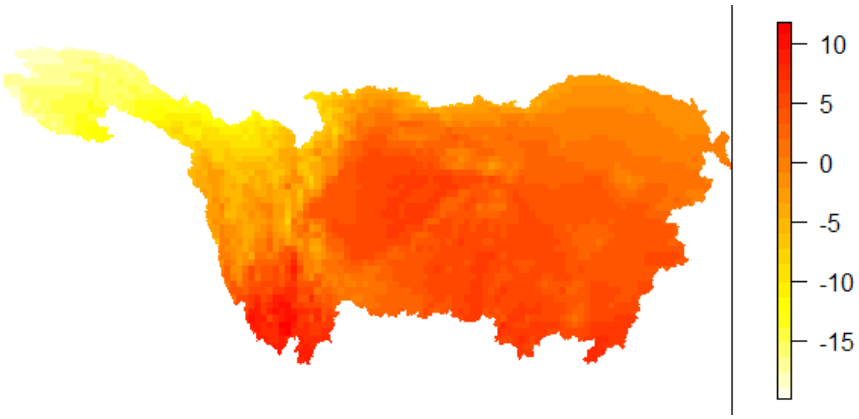


Figure 5.13: Temperature raster at original 0.25°x0.25° resolution [°C]

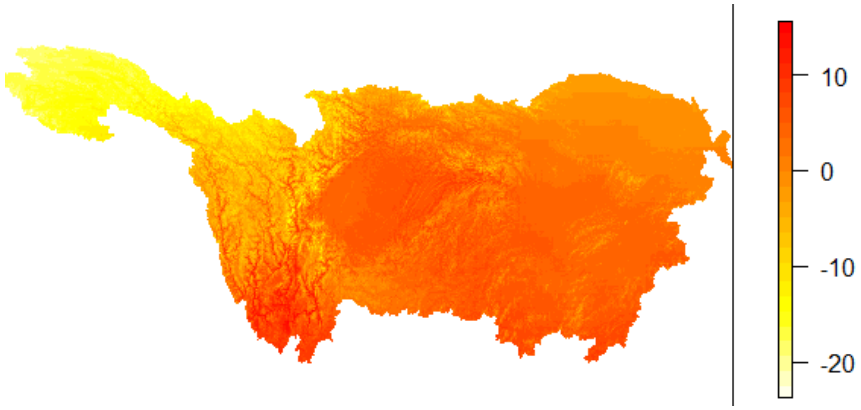


Figure 5.14: Final temperature raster at 15-arcsecond resolution [°C]

Although snowmelt and accumulation are dealt with in the *snowmodel* function, the *oneday* function attempts to track surface runoff freezing and melting, where hourly ice-melt is determined by:

$$\textbf{Hourly ice melt } [\textbf{mmh}^{-1}] = \frac{\textbf{DDF}_{\textbf{ice}} * (\textbf{Air temp} - \textbf{Melt temp})}{\textbf{24}} \quad (5.22)$$

where DDF_{ice} is the degree day factor of ice, set to 7.1 [mm of water $°C^{-1}$ day^{-1}], Melt temp is the ice melt temperature, assumed to be 1.5 [°C] and air temp is the air temperature.

A certain amount of surface runoff will refreeze if the temperature is below 0°C, determined by the temperature and the refreeze coefficient (R_{ice}) (Haeberli and Whiteman, 2015):

$$\textbf{Hourly ice refreeze } [\textbf{mmh}^{-1}] = \frac{\textbf{R}_{\textbf{ice}}\textbf{DDF}_{\textbf{ice}}\textbf{x}(\textbf{Air temp} - \textbf{0°C})}{\textbf{24}} \quad (5.23)$$

where R_{ice} is the refreeze coefficient of ice [0.05 mm of water $°C^{-1}$ day^{-1}]. Finally, the *oneday* function calls the *hourbyhour* function.

5.6.10 Running the model hour by hour

The *hourbyhour* function initiates the *runoffproduction* function at the start of each hour of the day, starting with time "00" and finishing with time "23". At the end of each day, the daily data and daily flow raster are saved to file and the daily flow image is plotted to the screen (if required).

5.6.11 Calculating runoff production

The *runoffproduction* function determines the amount of runoff produced in one hour calculated by the equations of the G2G model (see section 5.4). Hence all the runoff produced

is estimated at the beginning of one hour and then routed through the catchment during each timestep. The form of the precipitation falling as rain or snow is determined and separated, where rainfall is added directly to runoff in cells absent of lying snow together with any melted frozen runoff. Any runoff re-frozen is subtracted from the available runoff. If snow is lying anywhere within the catchment, or snow is falling, then the *snowmodel* function is called. The rest of the function calculates the new cell storage considering evapotranspiration, drainage from the cell and any runoff entering the cell storage which is dependent on the critical moisture value. Remaining runoff is sent to the *routing* function.

5.6.12 The Snow model

The *Snowmodel* function implements the SNOW-17 snow accumulation and ablation model (see section 4.5), where new snowfall is added to the catchment snow, and temperature and rainfall contribute to snow melt. Melt water is added to surface runoff.

5.6.13 *Routing* flow through the catchment

Routing simply calls the *route* function which is replicated several times per hour (the replication value). If desired, the hourly runoff is plotted to screen via the *plothourlyrunoff* function.

The *route* function tracks the time of the hour, day and total since starting by adding the timestep to the time thus far. Return flow is added to runoff and subtracted from baseflow, where:

$$\textbf{Return flow} = \textbf{Return flow rate constant x Baseflow} \qquad (5.24)$$

The runoff moved to the next cell and remaining runoff at the end of the timestep are calculated by:

$$\textbf{Remaining runoff} = \textbf{Runoff x Remaining Surface Rate constant} \qquad (5.25)$$

$$\textbf{Moved runoff} = \textbf{Runoff x Rate constant} \qquad (5.26)$$

A similar method is used to find moved and remaining baseflow. The moved runoff and baseflow is routed to the next cell via the *EastC, WestC* and *MoveC* Rccp functions described

earlier. Any moved runoff is then added to the remaining runoff and the hourly, daily and total runoff is calculated for each cell.

5.6.14 Plotting hourly runoff

If the *plothourlyrunoffswitch* is set to 1, then the *plothourlyrunoff* function is initiated. This calculates the runoff in ls^{-1} over that hour by:

$$\textbf{Hourly runoff } [\textbf{ls}^{-\textbf{1}}] = \frac{\textbf{hourly runoff} * \textbf{cell area} * \textbf{1000000}}{\textbf{3600}} \qquad (5.27)$$

where hourly runoff is the depth of surface water for each cell accumulated over one hour [mm] and cell area is the area of each raster cell [km^2]. Similar calculations are used to find the mean surface runoff and base flow for the whole day and the mean flow since the beginning of the model run. Flow values at the catchment outlet (or observation point) are extracted. By converting the hourly runoff matrix back to a raster, it can be plotted to screen along with other relevant statistics (see figure 5.15). Note this is only for model runs on a Windows machine.

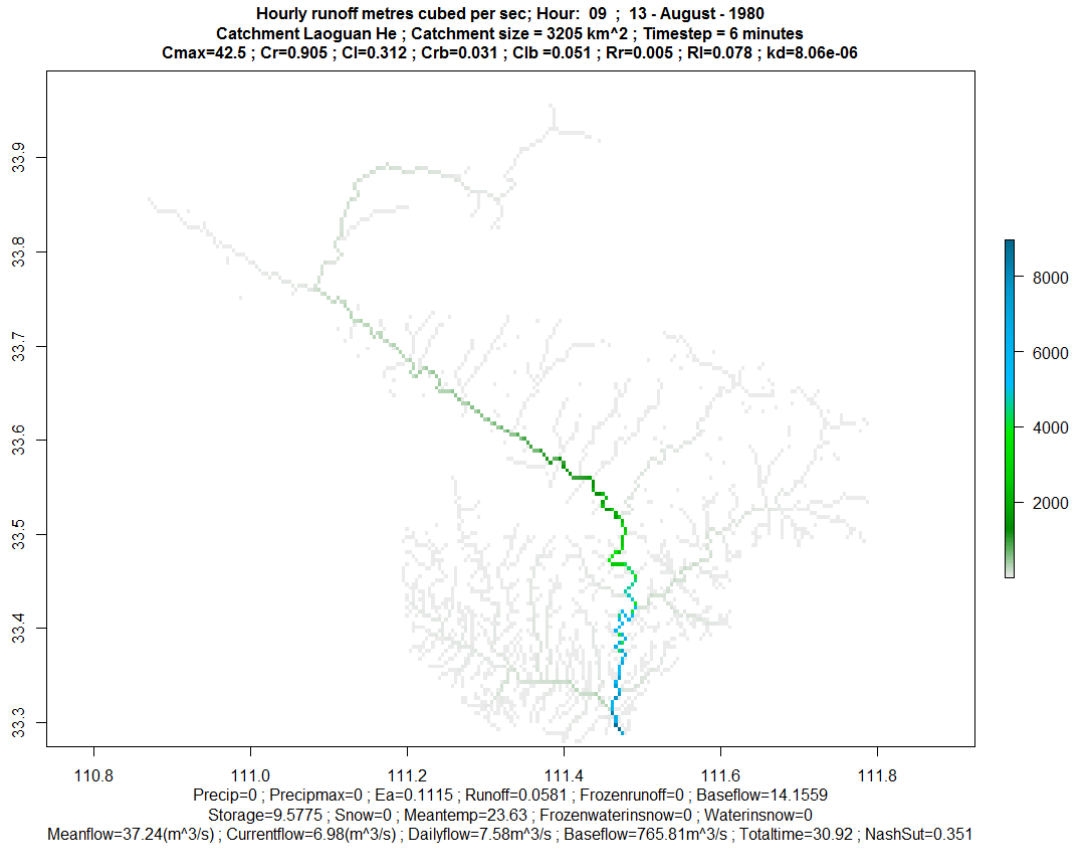


Figure 5.15: Example screen capture of the hourly runoff plot [ls^{-1}] for the Laoguan catchment, showing largest rivers in blue, moving to smallest rivers in grey. Blank areas are land or out of catchment

5.6.15 Saving data at the end of each day

At the end of each day, the *savedailydata* function captures the date and daily statistics from the catchment outlet (or observation point) and writes the data to file. The *Savedailyflowraster* function converts the mean daily flow matrix to a raster and writes it to file with a title of the day, month and year.

5.6.16 Plotting the daily flow to the screen

At the end of each simulated day, the *dailyflowplottoscreen* function plots the modelled mean daily surface runoff [m^3s^{-1}] at the catchment outlet / observation point (red) against the observed daily surface runoff (green) over the time-series of the model run (see figure 5.16).

The plot is accompanied by statistics on how the model compares to the different calibration objective functions (efficiency criteria) and the volume of runoff error (see section 5.8).

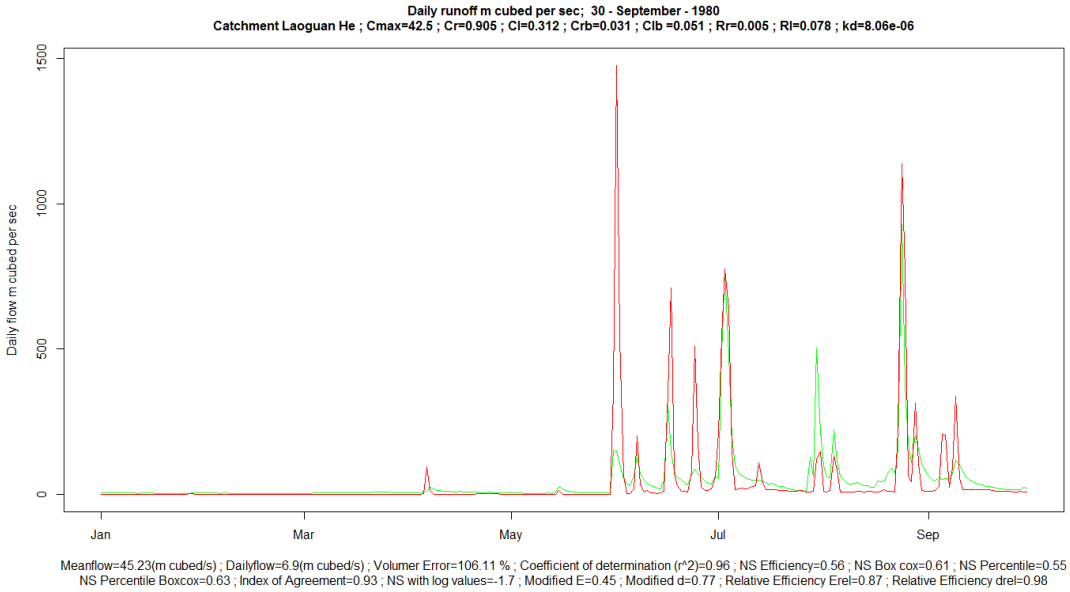


Figure 5.16: Example screen capture of the modelled (red) vs observed (green) mean daily surface runoff [m^3s^{-1}] for the Laoguan catchment with calibration performance for various objective functions

5.6.17 Saving data at the end of the model run

At the very end of the run, a plot similar to that of the *dailyflowplottoscreen* is captured and written to file as a PDF file (note that headless systems such as the EDDIE compute cluster cannot handle jpeg type capture). A flow duration curve of the plot is also captured by ordering the observed and modelled data in descending order, and capturing percentile values from 100th (largest flow) to 0th (smallest flow) percentile and plotting along with the calibration data (see figure 5.17).

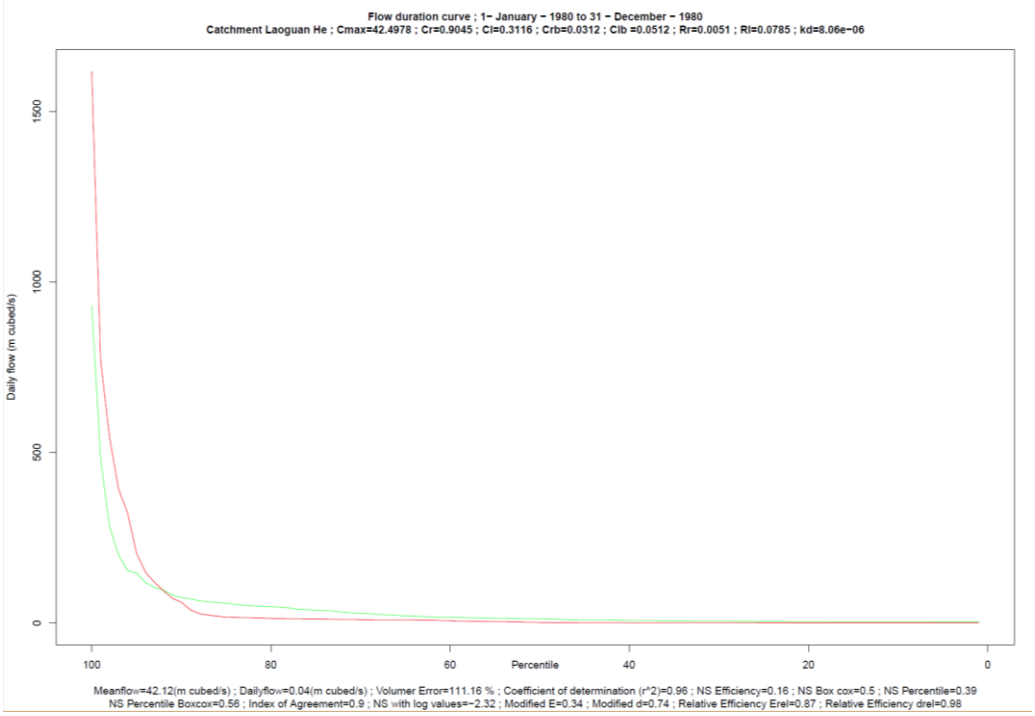


Figure 5.17: Example Flow duration curve (FDC) comparing modelled (red) and green (observed) surface runoff [m^3s^{-1}] at the end of the model run and calibration performance

5.7 Model performance

Model performance depends largely on the size of the catchment being modelled. The Laoguan He Xixia catchment (a sub basin of the YDB) at an estimated 3205km^2 takes approximately 45 minutes to model one whole year at 15-arcseconds on the Windows machine. Note that this small sub-catchment relative to the YDB is still comparable in size to the River Spey catchment, the third longest river in Scotland. Modelling of one year of the whole YDB at nearly 2 million km^2 with a 6-minute timestep takes approximately 14 days on the PC described. This highlights the difficulties in increasing resolution for such large catchments.

5.8 Calibration of the model – efficiency criteria

As introduced earlier, evaluation of hydrological model performance requires comparison of observed and modelled data, usually at the catchment outlet or at least somewhere down stream within the catchment. Visual inspection of the simulated and observed hydrographs is sometimes used where a hydrologist makes a subjective assessment of the model performance.

However, assessing the model 'fit' to simulated behaviour usually relies on comparison to an objective function or efficiency criteria (referred to in tables as EC) to provide an objective assessment. Efficiency criteria such as the Nash-Sutcliffe index are commonly used in hydrological modelling, but there are other criteria which place different emphasis on different types of simulated and observed behaviours.

Efficiency criteria are mathematical measures of how well a model fits available observations (Beven, 2001a) and many contain a summation of the error-term (i.e. the difference between the observed and simulated value at each timestep) normalised by a measure of the variability in the observations. The absolute or squared value is often used to avoid cancelling of errors, placing emphasis on larger errors. In hydrology terms, larger errors are usually associated with high stream-flows and calibration can lead to model parameters that fit peak flows at the expense of low flow. Therefore, introducing efficiency criteria that prioritise other attributes of the hydrograph is desirable.

In this model, a number of efficiency criteria (for further details see Krause et al., 2005) were included to ensure parameter sets are selected that give an overall strong performance against all the criteria, rather than just performing well against one. All the efficiency criteria used in this model have a maximum and optimum value of 1. A value of 0 or less usually indicates that the mean of the observed data would be a better predictor than the model itself. Catchments with efficiency criteria of 0.5 and above are considered behavioural.

Timing errors of meteorological data and stream-flow may lead to large errors between observed and modelled daily surface runoff, and therefore the efficiency criteria in this model are compared to the FDC percentile values instead of the observed and modelled flows at each timestep.

5.8.1 Coefficient of determination r^2

The coefficient of determination (r^2) is the squared value of the coefficient of correlation according to Bravais-Pearson calculated as:

$$\mathbf{r^2} = \left(\frac{\sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{O_i} - \overline{\mathbf{O}})(\mathbf{P_i} - \overline{\mathbf{P}})}{\sqrt{\sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{O_i} - \overline{\mathbf{O}})^{\mathbf{2}}} \sqrt{\sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{P_i} - \overline{\mathbf{P}})^{\mathbf{2}}}} \right)^{\mathbf{2}} \qquad (5.28)$$

where O_i is observed runoff values at each FDC percentile [m^3s^{-1}], P_i is predicted runoff values at each FDC percentile [m^3s^{-1}], $\overline{O}$ is mean observed runoff [m^3s^{-1}] and $\overline{P}$ is mean predicted runoff [m^3s^{-1}]. This efficiency criterion estimates the combined dispersion against the single dispersion of the observed and predicted series, and a value of 1 means the dispersion of the prediction is equal to the dispersion of the observed values. However, a parameter set which systematically over-or under predicts all the time can also give a value of 1 even when the predictions are wrong.

5.8.2 Nash Sutcliffe Efficiency E

Nash Sutcliffe (1970) efficiency has a range of 1 to -∞ and is calculated by:

$$\mathbf{E = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2}} \qquad (5.29)$$

A disadvantage of the Nash Sutcliffe efficiency is that the differences of observed and predicted values are squared, hence large stream flow errors are strongly overestimated. This overestimates model performance during peak flows and underestimates low flow conditions.

5.8.3 Index of agreement d

The index of agreement, d, (Willmot 1984) was intended to overcome the insensitivity of E and r^2 to differences in the observed and predicted means and variances, defined as:

$$\mathbf{d = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(|P_i - \overline{O}| + |O_i - \overline{O}|)^2}} \qquad (5.30)$$

5.8.4 Modified forms of E and d

Modified forms of E and d are often included with the same purpose as to increase the sensitivity to lower values where:

$$\mathbf{Modified\ E_j = 1 - \frac{\sum_{i=1}^{n}|O_i - P_i|^j}{\sum_{i=1}^{n}|O_i - \overline{O}|^j}\ with\ j \in N} \qquad (5.31)$$

$$\mathbf{Modified\ d_j = 1 - \frac{\sum_{i=1}^{n}|O_i - P_i|^j}{\sum_{i=1}^{n}(|P_i - \overline{O}||O_i - \overline{O}|)^j}\ with\ j \in N} \qquad (5.32)$$

with j = 1, the overestimation of flood peaks is reduced significantly resulting in better overall evaluation.

5.8.5 Relative efficiency criteria Erel and drel

All the criteria discussed so far quantify the absolute values of the difference between observation and prediction, giving a greater influence on higher values of over-or under prediction. Efficiency criteria based on relative deviations are used to counter-act this:

$$\mathbf{E_{rel}} = \mathbf{1} - \frac{\sum_{\mathbf{i=1}}^{\mathbf{n}} \left(\frac{\mathbf{O_i} - \mathbf{P_i}}{\mathbf{O_i}}\right)^{\mathbf{2}}}{\sum_{\mathbf{i=1}}^{\mathbf{n}} \left(\frac{\mathbf{O_i} - \mathbf{P_i}}{\mathbf{O_i}}\right)^{\mathbf{2}}} \qquad (5.33)$$

$$\mathbf{d_{rel}} = \mathbf{1} - \frac{\sum_{\mathbf{i=1}}^{\mathbf{n}} \left(\frac{\mathbf{O_i} - \mathbf{P_i}}{\mathbf{O_i}}\right)^{\mathbf{2}}}{\sum_{\mathbf{i=1}}^{\mathbf{n}} \left(\frac{|\mathbf{P_i} - \bar{\mathbf{O}}| + |\mathbf{O_i} - \bar{\mathbf{O}}|}{\mathbf{O_i}}\right)^{\mathbf{2}}} \qquad (5.34)$$

This reduces the influence of absolute differences during high flows significantly, and the influence of absolute lower flow differences are enhanced.

5.8.6 Box-Cox transform of Nash-Sutcliffe (E_{BoxCox})

Box-Cox (1964) transforms are commonly applied to data to reduce the skewness in error across the flow range where:

$$\mathbf{O_{iBoxCox}} = \frac{[(\mathbf{O_i}+\mathbf{1})^{\boldsymbol{\gamma}}-\mathbf{1}]}{\boldsymbol{\gamma}} \qquad (5.35)$$

$$\mathbf{P_{iBoxCox}} = \frac{[(\mathbf{P_i}+\mathbf{1})^{\boldsymbol{\gamma}}-\mathbf{1}]}{\boldsymbol{\gamma}} \qquad (5.36)$$

where γ is a parameter typically set to 0.3 (Vrugt, 2003). These values are substituted into the Nash-Sutcliffe efficiency:

$$\mathbf{E_{BoxCox}} = \mathbf{1} - \frac{\sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{O_{iBoxCox}} - \mathbf{P_{iBoxCox}})^{\mathbf{2}}}{\sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{O_{iBoxCox}} - \bar{\mathbf{O}}_{\mathbf{BoxCox}})^{\mathbf{2}}} \qquad (5.37)$$

5.8.7 Multi-objective efficiency criteria

Krause et al. (2005) conclude that for sound scientific model calibration and validation, a combination of different efficiency criteria complemented by volume error is recommended, where volume error is computed as:

$$\textbf{Volume error} = \left(\frac{\textbf{Total volume of modelled flow } [\mathbf{m^3}]}{\textbf{Total volume of observed flow } [\mathbf{m^3}]} * \mathbf{100}\%\right) - \mathbf{100}\% \qquad (5.38)$$

A volume error of 0% would mean that the modelled and observed flow were equal. In this project, the eight efficiency criteria are calculated at the end of each model run, and the mean taken to produce an overall efficiency criterion with a maximum value of 1. The volume error is used to assess the suitability for hydropower search. A large volume error under – or overestimates the flow through a river and hence may overestimate the energy potential of a hydropower scheme. A conservative, underestimate is preferred in costing viable hydropower schemes.

In assessing model 'fit' it was decided to take the mean of the eight efficiency criteria described earlier i.e. r^2, E, d, Mod E, Mod d, E_{rel}, d_{rel} and E_{BoxCox}.

5.9 Selection of a parameter search space

With the efficiency criteria defined, a parameter calibration (or optimisation) methodology was then employed to minimise (or maximise) the efficiency criteria. As the model has 8 undefined parameters, completely random values would give no guarantee of successfully minimising the objective function. To increase the likelihood of finding parameter sets that give a good 'fit', the parameter space was constrained using previous studies employing the G2G model which found the following optimum parameters:

	Parameters							
Study	c_{max}	c_r	c_l	c_{rb}	c_{lb}	r_l	r_r	k_d
Duncan (2014)	194	1.5	1.1	0.15	0.14	0.096	0.43	$2*10^{-7}$
Duncan (2014)	200	1.18	1	0.09	0.06	0.09	0.18	$1*10^{-7}$
Duncan (2014)	162	1.34	1.1	0.14	0.13	0.173	0.81	$1*10^{-7}$
Cole and Moore (2008)	40	1.5	0.07	0.5	0.05	0.05	0.05	$1.5*10^{-6}$
Bell et al. (2007)	140	0.5	0.4	0.05	0.05	0.005	0.005	$5*10^{-5}$

Table 5.4: Optimum parameters from other studies using the G2G model

Based on the above, the final parameter space was defined as:

	Parameters							
	c_{max}	c_r	c_l	c_{rb}	c_{lb}	r_l	r_r	k_d
Lower bound	40	0.1	0.1	0.01	0.01	0.001	0.001	$1*10^{-8}$
Upper bound	250	2.0	2.0	0.5	0.3	0.2	1.0	$1*10^{-5}$

Table 5.5: Final defined parameter space for optimisation algorithm

5.10 Selection of a parameter optimisation algorithm

With 8 parameters to calibrate, even a 10-step sweep across each parameter would require 10^8 model runs, clearly unfeasible for this model with run times of minutes, hours or days. Monte Carlo methods are computational algorithms that rely on random sampling of the parameter space. Guided Monte Carlo methods include the Markov Chain Monte Carlo (MCMC) sampler (Hastings, 1970) and the METROPOLIS algorithm (Metropolis et al., 1953; Kukzera and Parent, 1998).

Automated calibration methodologies to optimise efficiency criteria include simulated annealing, genetic evolution, particle swarm optimisation, branch and bound methods, deterministic and other stochastic methods. There are many R packages that enable these methodologies, and these were reviewed by Mullen (2014). Of those packages, three enable parallelisation, the ability to run the model across multiple cores, hence enabling multiple model runs at the same time: *rgenoud* (Mebane and Sekhon, 2015), *DEoptim* (Ardia et al., 2016) and *ppso* (Franke 2012). *DEoptim* was selected due to it consistently minimalising efficiency criteria compared to other packages in most of the tests with varying number of parameters and its ability to run error free (*ppso* and *rgenound* both reported errors across several test runs).

DEoptim implements the differential evolution algorithm (Storn and Price, 1997) for global optimisation of a real-valued function of a real-valued parameter vector. It is an optimisation algorithm inspired by the process of natural selection, an evolutionary algorithm, performing crossover, mutation and selection on a parameter-set population. Genetic algorithms are proven heuristic methods for global optimisation, particularly for combinatorial optimisation problems (Ardia et al., 2016).

At each generation, the algorithm transforms a set of parameter vectors (termed the population) into another set that are more likely to minimise the objective function, by disturbing an old parameter vector with the scaled difference of two randomly selected parameter vectors. At generation 0, guesses are made for the optimal values of the parameter vector using random values between the upper and lower bounds. Each generation involves creation of a new population from the current population members, accomplished by differential mutation. Mutation is continued, and a crossover probability controls the fraction of parameter values copied over from the mutant. If a trial parameter vector generates an equal or lower objective function, this replaces the original parameter set of the population. The algorithm terminates after a set number of generations, if a threshold objective function is reached or it is unable to reduce the objective function over a set number of generations.

5.11 The calibration strategy

The optimisation algorithm selected (*DEoptim*) relies on multiple model runs, preferably hundreds if not thousands to adequately sample the parameter space. The model runtimes for the whole YDB are in the region of two weeks to represent one year, and even a few model runs would take longer than is available. Therefore, a strategy was devised to calibrate sub-catchments of the YDB with view to finding an optimum parameter set to run across the whole of the YDB.

As introduced in section 3.4.9, the GRDC published catchment shapefile and river flow data for a number of catchments across China, several of which are within the YDB and listed in table 5.6 (where 'He' is Chinese for river) and depicted in figure 5.18:

River name	Gauging Station	HydroSHEDS Catchment Area (km^2)	Daily Data available	Missing data %
Bailu He	Baiqueyua	290	1977 to 1996	5.01%
Zagunao He	Zagunao	2,400	1981 to 2000	14.99%
Huai He	Changtaigua	3,025	1974 to 1993	2.1%
Laoguan He	Xixia	3,216	1977 to 1996	10.01%
Yalong Jiang	Luning	107,882	1981 to 2000	0%
Tongtian He	Zhimenda	134,399	1978 to 1997	0%
Yangtze River	Datong	1,679,569	2004	0%

Table 5.6: GRDC river flow data availability of YDB sub-catchments for calibration

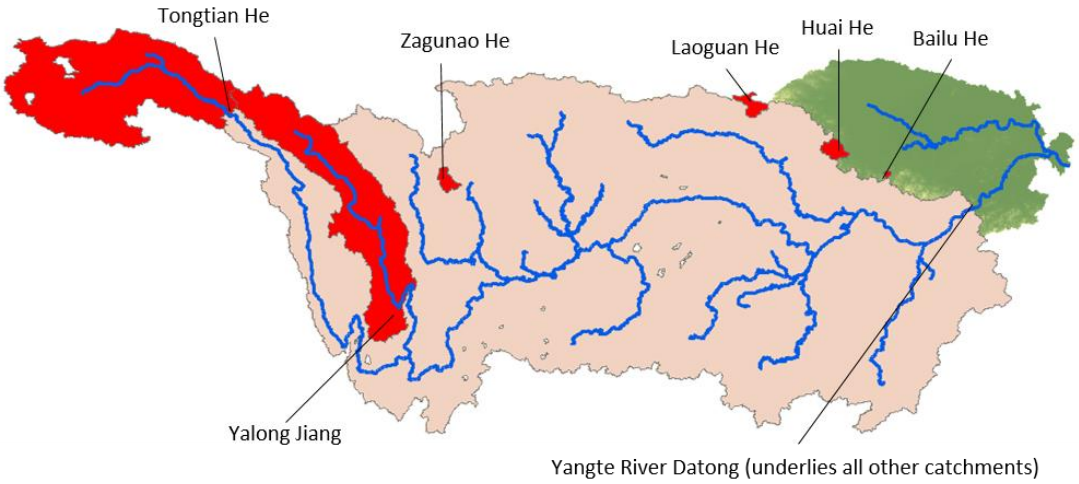


Figure 5.18: Representation of calibration catchments (red) within the YDB and the validation Yangtze River Datong catchment (salmon colour)

Of the catchments listed above, only four are small enough to perform sufficient model runs to enable the differential evolution algorithm to perform adequately within the timeframe i.e. Bailu He, Zagunao He, Huai He and Laoguan He. Therefore, the following calibration strategy was derived:

1. The differential evolution algorithm was performed on each of the four small catchments for one year only (e.g. January to December,1979).
2. Any strong performing parameter sets (e.g. average efficiency criteria values over 0.8) were tested over a three-year period on the same small catchments (e.g. January 1979 to December 1982).
3. The best 10 performing parameter sets from step 2 were tested over the entire period of the calibration data for each small catchment (e.g. January 1979 to December 1996).
4. The best performing parameter set from each of the small catchment calibrations (step 3) was trialled on the three other small catchments, plus the two larger catchments Yalong Jiang and Tongtian He in a round robin type test i.e. four parameter sets each tested against six small and medium sized catchments.
5. The best overall performing parameter set was validated across the six catchments on the large Yangtze River Datong catchment.

Ideally the model would be set with different parameters across different sub-catchments, and as the model is a distributed model, this is to some extent possible. However, the model relies

on having a consisting timestep for each cell and hence a constant wave-speed is required (see equation 5.13) (i.e. the max of c_r and c_l needs to be constant across any catchment modelled). Hence it was decided to select one parameter set for the whole catchment.

5.12 Calibration results

5.12.1 Bailu He (Baiqueyua) Catchment Calibration

The Bailu He Baiqueyua catchment is a small 290km^2 sub-catchment in the north-east of the YDB with elevations ranging from 77m to 826m. Calibration of the year 1979 yielded mean efficiency criteria results ranging from -9.265 to 0.97 (maximum of 1). Out of nearly 5000 model runs, 1303 parameter sets resulted in a mean efficiency criterion result of 0.8 and over, and these were taken forward to the next round of 3-year model runs. The best performing parameter set FDC of round 1 is shown in figure 5.19 and the parameters and efficiency criteria shown in table 5.7.

Catchment	**C_{max}**	**c_r**	**c_l**	**c_{rb}**	**c_{lb}**	**R_r**	**R_l**	**k_d**
Bailu He, Baiqueyua 1979	109.747	0.608	0.179	0.244	0.173	0.145	0.198	$1.01*10^{-6}$
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
+0.69%	1.00	0.99	1.0	0.93	0.96	0.96	0.99	0.99

Table 5.7: Best performing parameter set and efficiency criteria results from one year calibration of the Bailu He Baiqueyua catchment over 1979 (mean EC = 0.98)

Although the parameters within table 5.7 give the best mean efficiency criteria, there were many other quite distinct parameter sets which also gave good model fits found across the range of each parameter. This is known as equifinality and table 5.8 shows the range of parameter values for those models with a mean efficiency criterion result of 0.9 or over (i.e. a good model fit).

Parameter	Parameter limits	Good model fit parameter range where mean efficiency criteria > 0.9
C_{max}	40 – 250	50.312 – 232.420 [mm]
c_r	0.1 – 2.0	0.105 -1.924 [ms^{-1}]
c_l	0.1 -2.0	0.106 – 1.77 [ms^{-1}]
c_{rb}	0.01 – 0.5	0.007 – 0.447 [ms^{-1}]
c_{lb}	0.01 – 0.3	0.003 – 0.297 [ms^{-1}]
R_r	0.001 – 1.0	0.007 – 0.979
R_l	0.001 – 0.2	0.002 – 0.197
k_d	$1*10^{-8}$ – $1*10^{-5}$	$9.12*10^{-7}$ to $9.95*10^{-6}$

Table 5.8: Parameter ranges resulting in a good model fit where the average efficiency criteria is > 0.9

Table 5.8 highlights how good model fits can be achieved across the calibration limits for each parameter, either suggesting the parameter limits can be opened wider or the parameters are not particularly sensitive, and it is the combination of parameter sets that is more important. Figure 5.19 shows how the modelled results generally fit the FDC of the observed results.

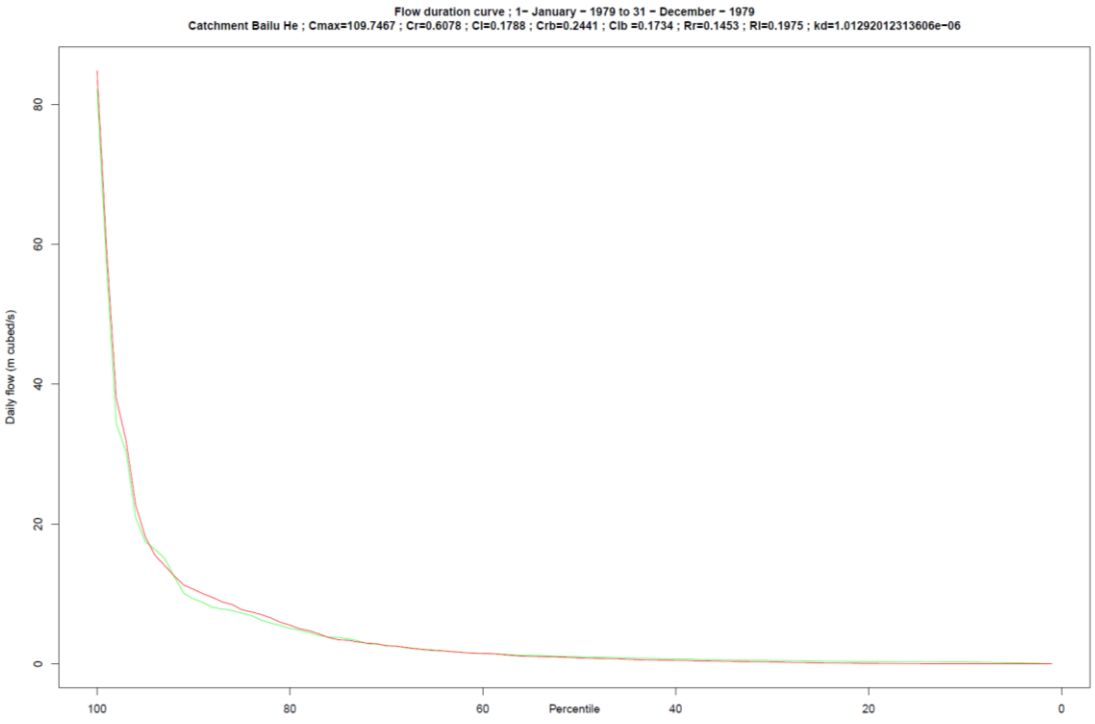


Figure 5.19: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Bailu He Baiqueyua, 1979

During the 3-year calibration, the best performing parameter set is shown in table 5.9 and the modelled vs observed FDC is shown in figure 5.20. Once again, this shows a good fit over the range of Q_{100} (minimum flow) to Q_5, but underestimates the peak flood flows.

Calibration	**C_{max}**	**c_r**	**c_l**	**c_{rb}**	**c_{lb}**	**R_r**	**R_l**	**k_d**
Bailu He, Baiqueyua 1979-1981	124.247	0.390	0.582	0.002	0.175	0.078	0.101	$3.28*10^{-7}$
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
-10.63	0.82	0.73	0.89	0.84	0.92	0.96	0.99	0.97

Table 5.9: Best performing parameter set and efficiency criteria results from three-year calibration of the Bailu He Baiqueyua catchment over 1979-1981 (mean EC = 0.89)

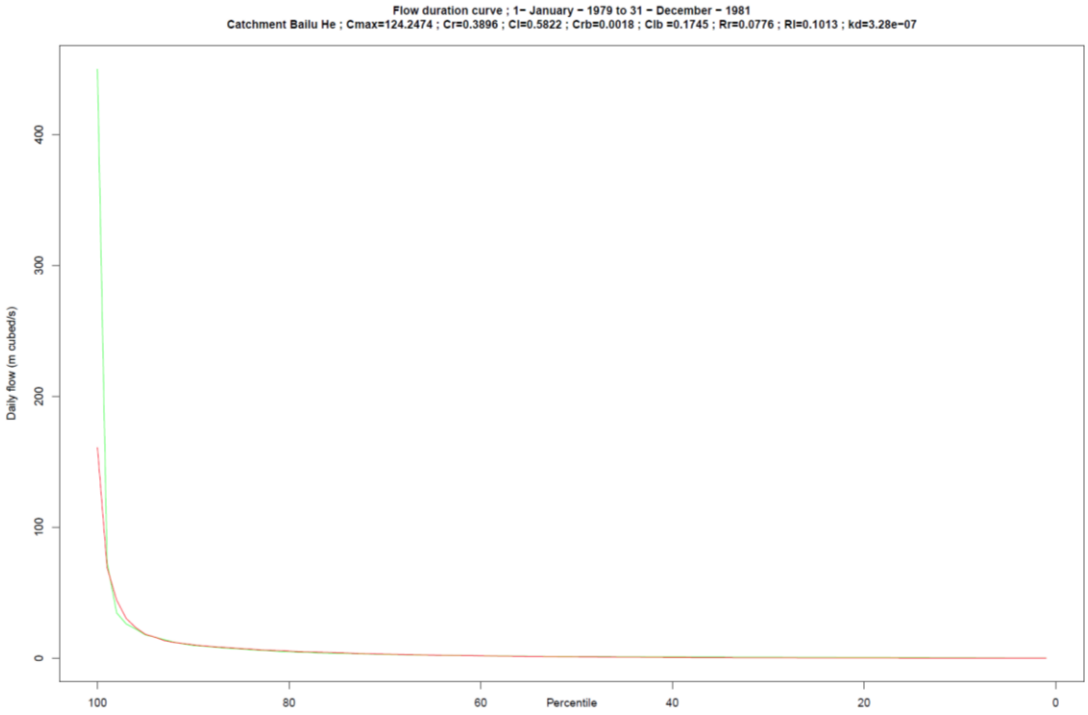


Figure 5.20: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Bailu He Baiqueyua, 1979 to 1981

The ten best performing parameter sets were modelled over the full length of the calibration data (1979 to 1996, 1992 missing), with the results shown in table 5.10 and the full efficiency criteria results of the best performing parameter set (set 7) is shown in table 5.11. The resulting FDC curve and mean daily flow plots are shown in figures 5.21 and figure 5.22 respectively,

again performing well except for the flood flows. This parameter set was selected for the round robin test across other sub-catchments.

Set	C_{max}	c_r	c_l	c_{rb}	c_{lb}	R_r	R_l	k_d	Mean Efficiency Criteria Bailu He 1979 to 1996
1	78.154	0.622	0.639	0.269	0.296	0.087	0.072	$1.03*10^{-6}$	0.91
2	87.587	0.852	0.991	0.414	0.084	0.097	0.093	$1.23*10^{-6}$	0.94
3	93.655	0.968	0.360	0.020	0.116	0.282	0.091	$3.36*10^{-7}$	0.94
4	94.463	0.830	0.873	0.210	0.299	0.061	0.176	$1.33*10^{-6}$	0.93
5	95.954	0.446	1.848	0.103	0.169	0.116	0.096	$7.75*10^{-7}$	0.94
6	106.167	1.251	0.573	0.150	0.008	0.800	0.115	$2.83*10^{-7}$	0.93
7	120.577	0.256	0.883	0.066	0.039	0.448	0.150	$1.73*10^{-7}$	0.95
8	124.247	0.390	0.582	0.002	0.175	0.078	0.101	$3.28*10^{-7}$	0.91
9	124.584	0.185	0.582	0.208	0.099	0.078	0.101	$3.28*10^{-7}$	0.91
10	138.416	0.272	0.272	0.079	0.249	0.860	0.121	$7.83*10^{-8}$	0.91

Table 5.10: Ten best performing parameter sets of the 3-year calibration of Bailu He Baiqueyua 1979 to 1981 and the mean efficiency criteria when modelling 1979 to 1996

Calibration	C_{max}	c_r	c_l	c_{rb}	c_{lb}	R_r	R_l	k_d
Bailu He, Baiqueyua 1979-1996	120.577	0.256	0.883	0.066	0.039	0.448	0.150	$1.73*10^{-7}$
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
-14.52	0.94	0.9	0.97	0.88	0.94	0.99	1	0.98

Table 5.11: Best performing parameter set and efficiency criteria results from calibration of the Bailu He Baiqueyua over 1979-1996 (mean EC = 0.95)

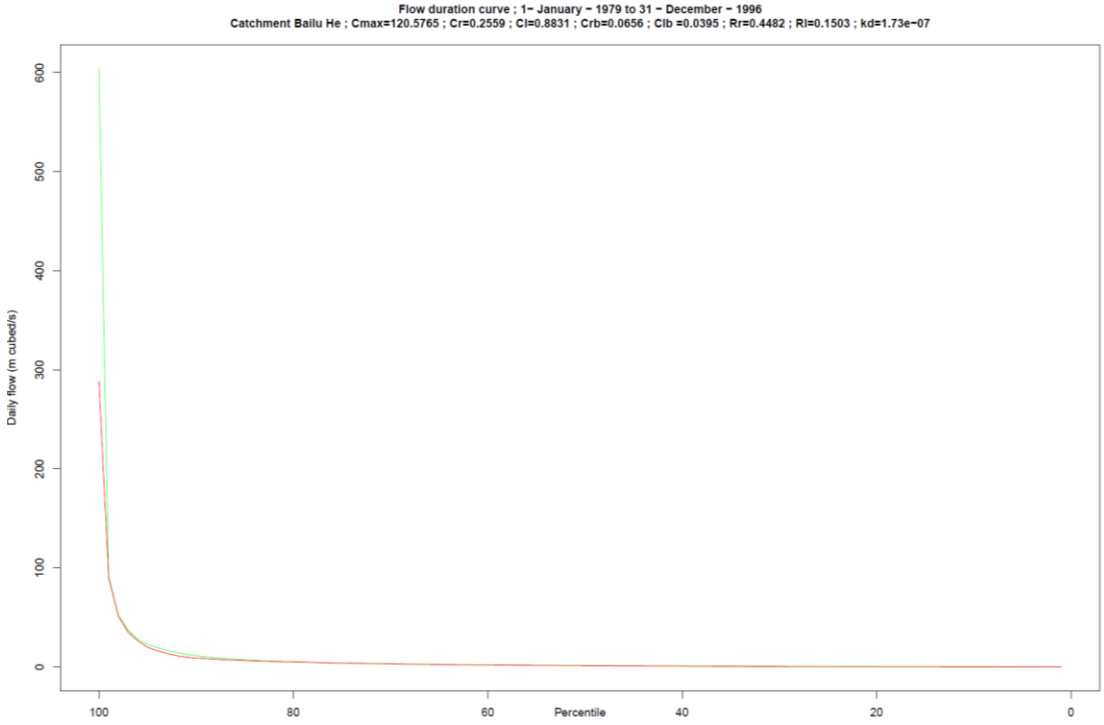


Figure 5.21: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Bailu He Baiqueyua, 1979 to 1996

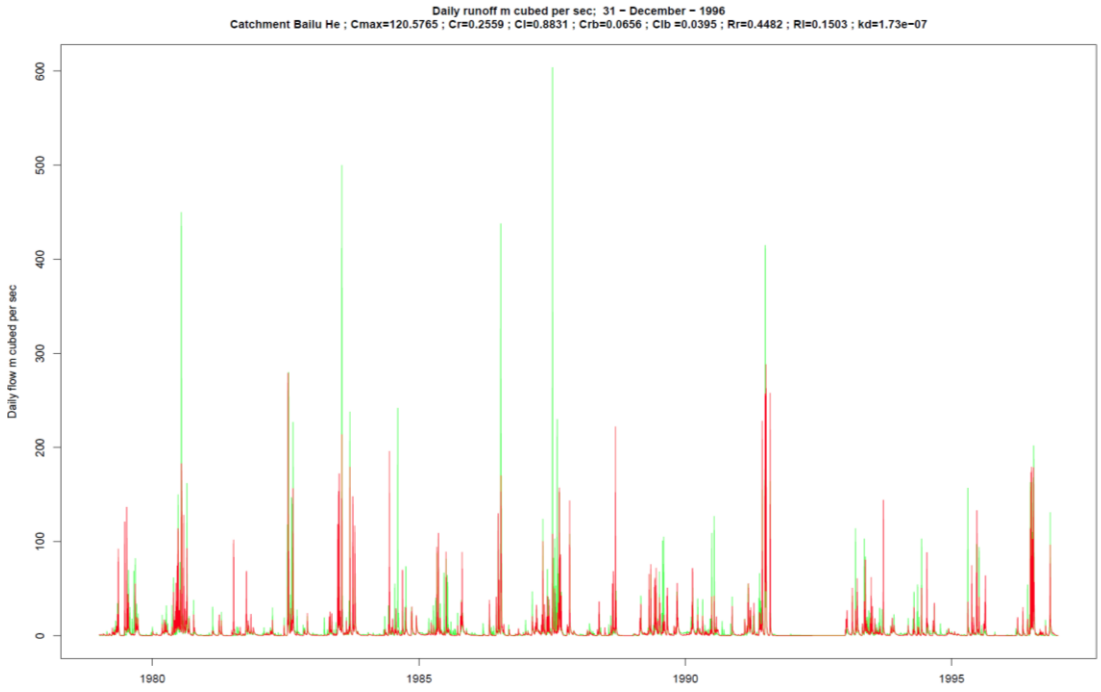


Figure 5.22: Modelled (red) and observed (green) mean daily flow [m^3s^{-1}] of the best performing parameter set during calibration of Bailu He Baiqueyua, 1979 to 1996

5.12.2 Laoguan He (Xixia) Catchment Calibration

The Laoguan He (Xixia) sub-catchment is a 3,216km^2 head catchment (i.e. its boundaries are on the YDB boundaries) ranging in height from 214m to 2018m. Following the methodology of the Bailu He Baiqueyua sub-catchment calibration, tables 5.12, 5.13 and 5.14 show the best performing parameter set and associated efficiency criteria for the one year (1980), three years (1979 to 1981) and full dataset calibrations (1979 to 1996: note 1993 and 1994 missing) respectively. Figures 5.23, 5.24 and 5.25 show the associated FDCs and figure 5.26 is the mean daily flow over the full calibration record.

Calibration	**C_{max}**	**c_r**	**c_l**	**c_{rb}**	**c_{lb}**	**R_r**	**R_l**	**k_d**
Laoguan He, Xixia 1980	95.315	1.015	0.806	0.020	0.080	0.016	0.003	1.64*10^{-6}
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
+2.19	0.98	0.97	0.99	0.81	0.91	0.93	0.98	0.86

Table 5.12: Best performing parameter set and efficiency criteria results from calibration of the Laoguan He Xixia catchment over 1980 (mean EC = 0.93)

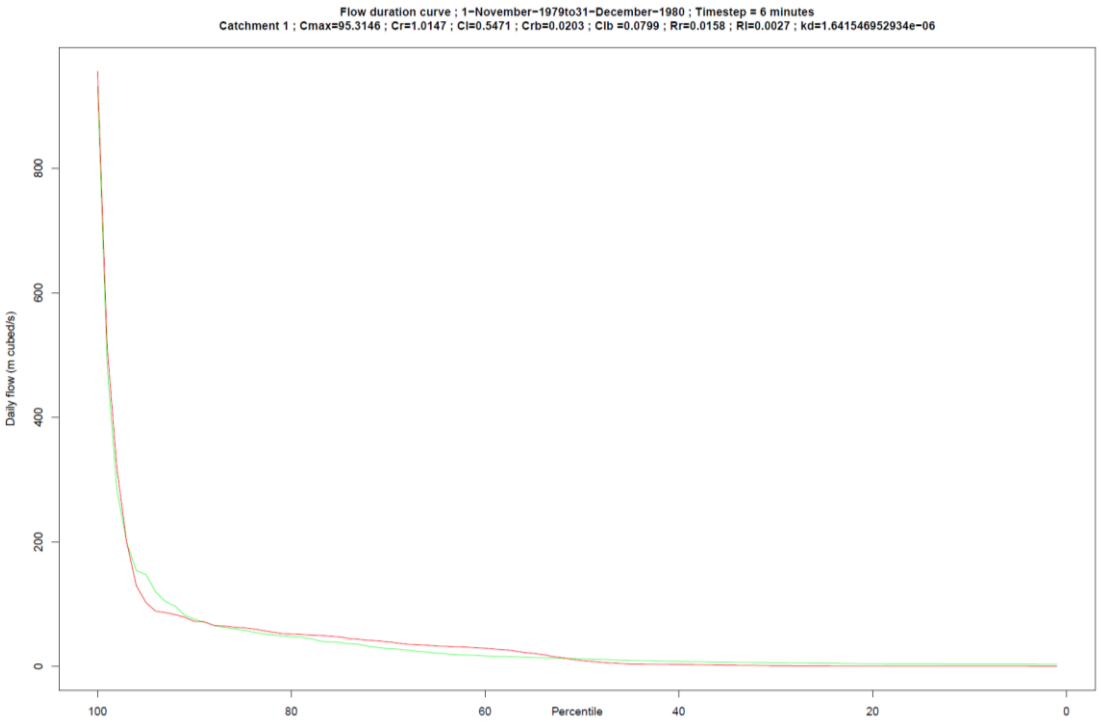


Figure 5.23: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Laoguan He Xixia, 1980

Calibration	C_{max}	c_r	c_l	c_{rb}	c_{lb}	R_r	R_l	k_d
Laoguan He, Xixia 1979-1981	89.920	1.487	0.967	0.027	0.080	0.012	0.049	$2.94*10^{-6}$
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
+9.91	0.96	0.94	0.99	0.76	0.89	0.93	0.99	0.86

Table 5.13: Best performing parameter set and efficiency criteria results from calibration of the Laoguan He Xixia catchment over 1979-1981 (mean EC = 0.92)

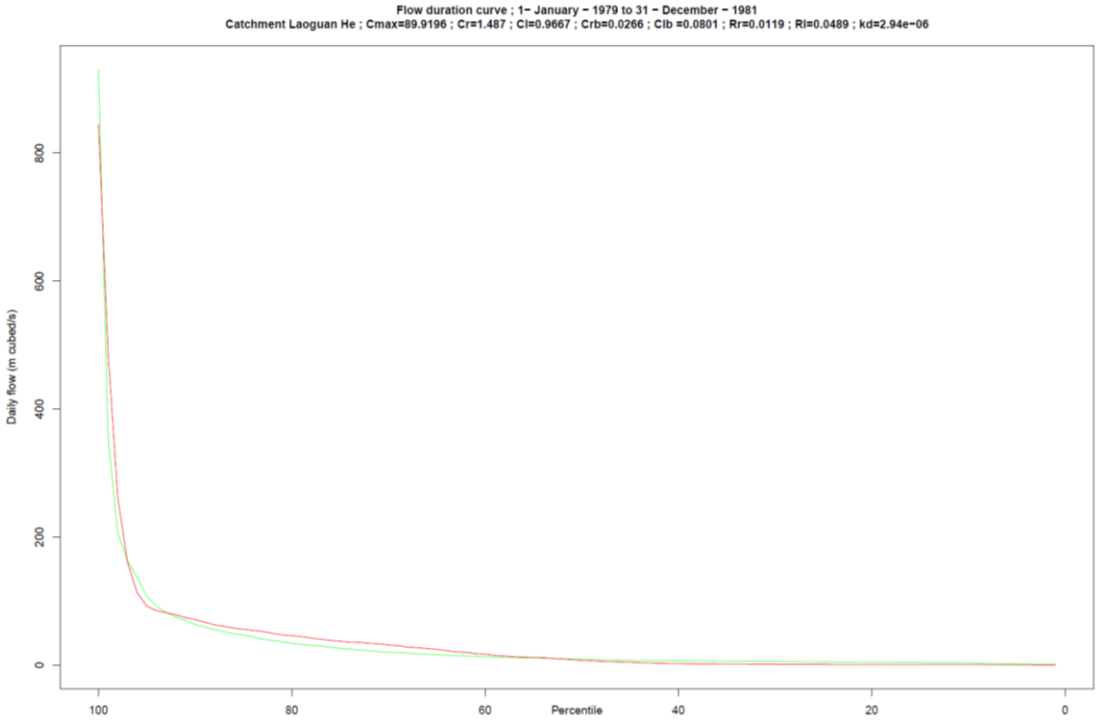


Figure 5.24: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Laoguan He Xixia, 1979 to 1981

Calibration	C_{max}	c_r	c_l	c_{rb}	c_{lb}	R_r	R_l	k_d
Laoguan He, Xixia 1979-1996	125.963	0.135	0.546	0.056	0.050	0.088	0.013	$4.53*10^{-7}$
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
+3.11	0.99	0.98	1.0	0.89	0.95	0.97	0.99	0.98

Table 5.14: Best performing parameter set and efficiency criteria results from calibration of the Laoguan He Xixia catchment over 1979-1996 (mean EC = 0.97)

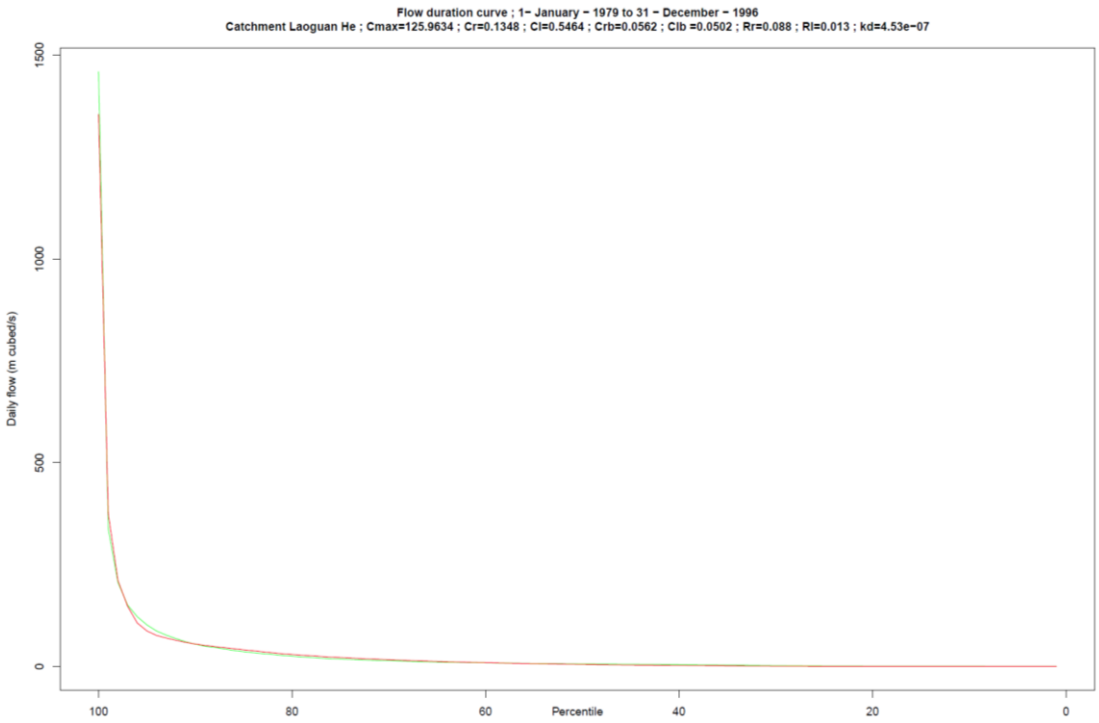


Figure 5.25: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Laoguan He Xixia, 1979 to 1996

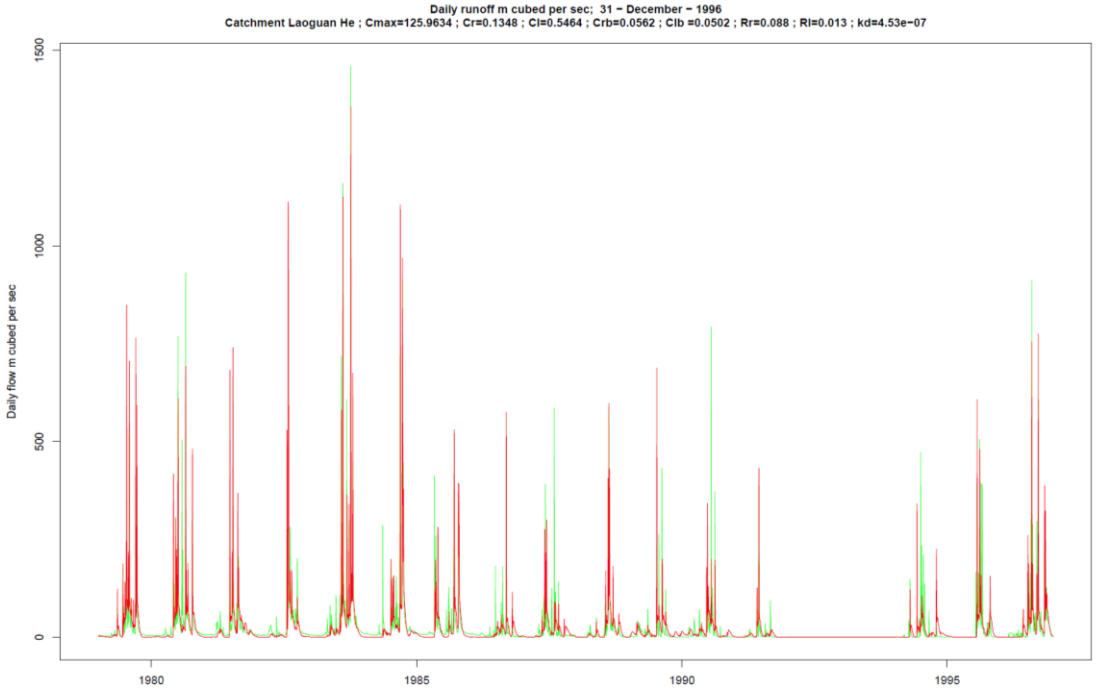


Figure 5.26: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Laoguan He Xixia, 1979 to 1996

There is generally good calibration over the full dataset with some underestimation of low-flow conditions with a resulting lower modified E result (0.89).

5.12.3 Huai He (Changtaigua) Catchment Calibration

The Huai He Changtaigua sub-catchment is 3,025km^2 with heights ranging from 47m to 990m. Again, following the methodology of calibration of the Bailu He and Laoguan He sub-catchments, the results for the best performing parameter set for one year (1980), three year (1979 to 1981) and the full length of the calibration data available (1979 to 1991) are given in tables 5.15, 5.16 and 5.17. The associated flow duration curves and full calibration daily flow are shown in figures 5.27, 5.28, 5.29 and 5.30.

Calibration	**C_{max}**	**c_r**	**c_l**	**c_{rb}**	**c_{lb}**	**R_r**	**R_l**	**k_d**
Huai He, Changtaigua 1980	94.296	1.110	0.597	0.012	0.088	0.006	0.094	5.86*10^{-7}
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
-5.75	0.97	0.96	0.99	0.82	0.91	0.96	0.99	0.86

Table 5.15: Best performing parameter set and efficiency criteria results from calibration of the Huai He Changtaigua catchment over 1980 (mean EC = 0.93)

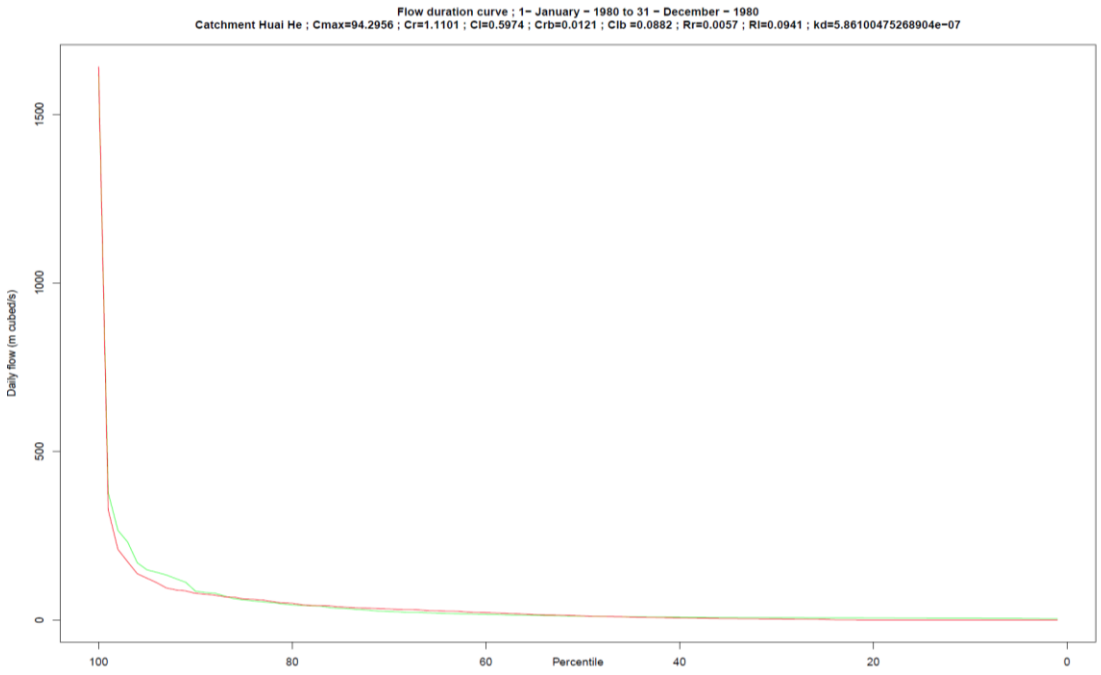


Figure 5.27: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Huai He Changtaigua, 1980

Calibration	C_{max}	c_r	c_l	c_{rb}	c_{lb}	R_r	R_l	k_d
Huai He Changtaigua 1979-1981	76.650	0.404	0.817	0.041	0.074	0.086	0.004	$9.58*10^{-7}$
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
-11.73	0.98	0.96	0.99	0.83	0.92	0.95	0.99	0.84

Table 5.16: Best performing parameter set and efficiency criteria results from calibration of the Huai He Changtaigua catchment over 1979-1981 (mean EC = 0.93)

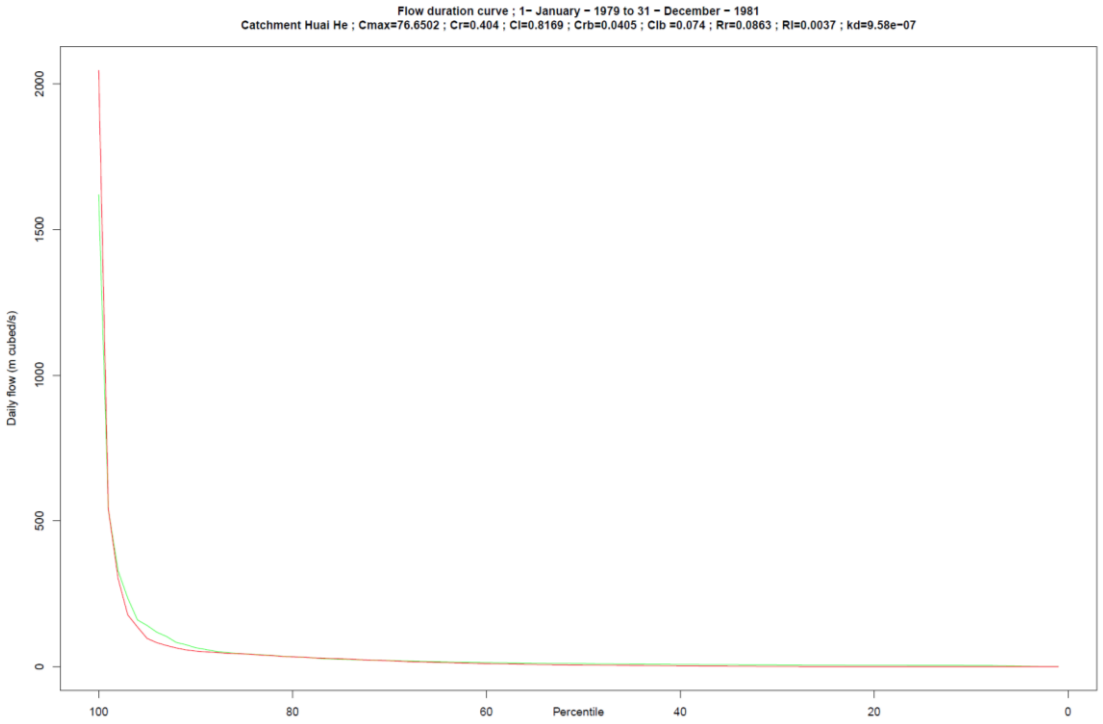


Figure 5.28: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Huai He Changtaigua, 1979 to 1981

Calibration	C_{max}	c_r	c_l	c_{rb}	c_{lb}	R_r	R_l	k_d
Huai He, Changtaigua 1979-1996	42.499	0.905	0.312	0.031	0.051	0.005	0.079	$8.06*10^{-6}$
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
+4.56	0.97	0.97	0.99	0.86	0.93	0.97	0.99	0.92

Table 5.17: Best performing parameter set and efficiency criteria results from calibration of the Huai He Changtaigua catchment over 1979-1996 (mean EC = 0.95)

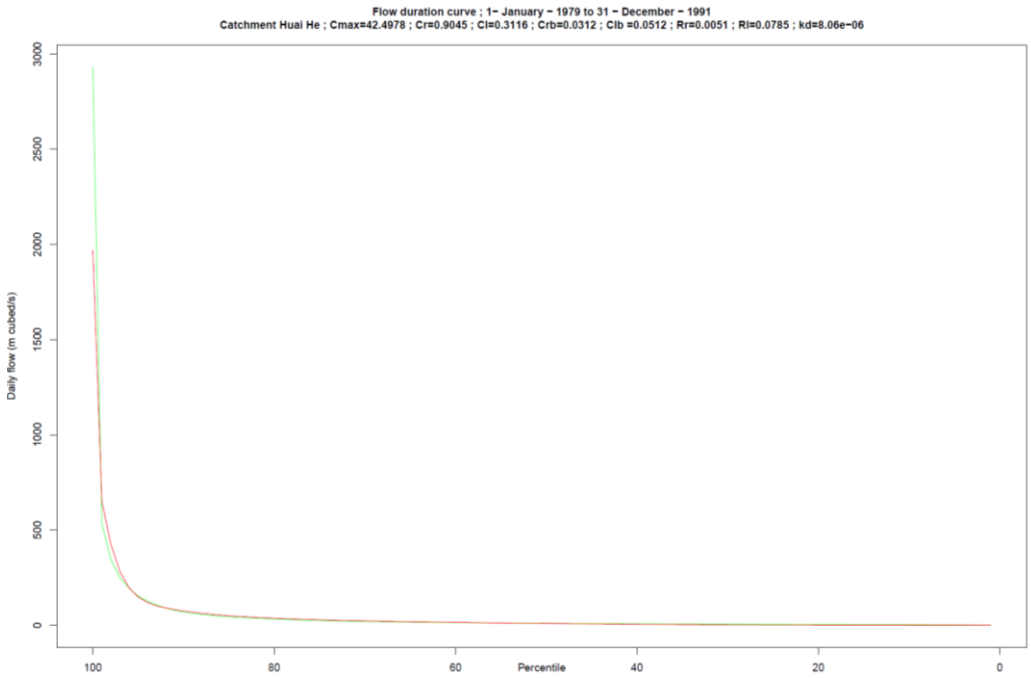


Figure 5.29: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Huai He Changtaigua, 1979 to 1991

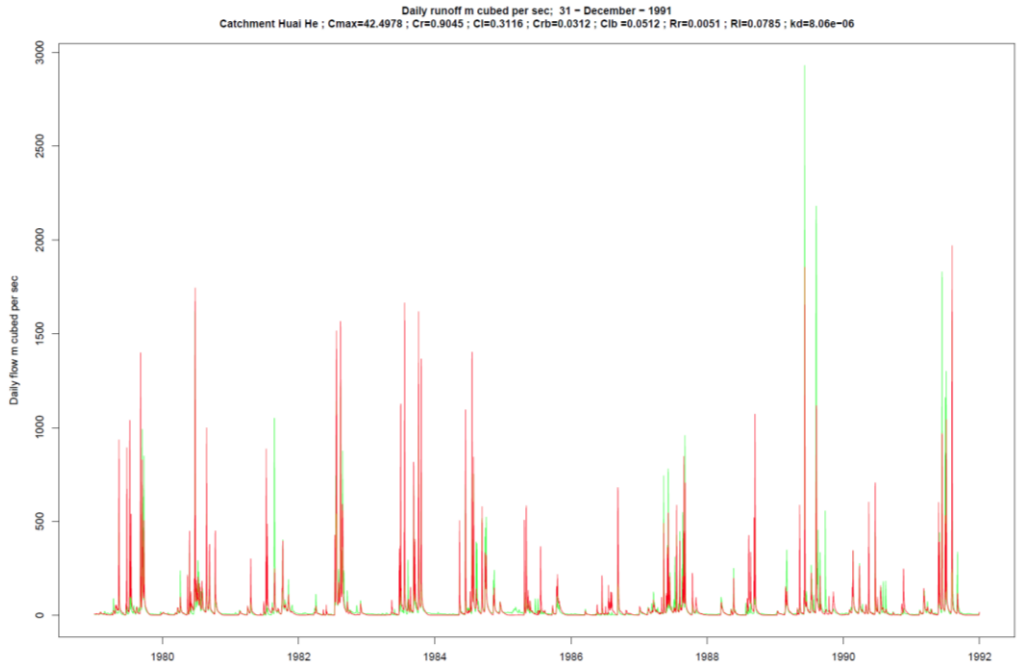


Figure 5.30: Modelled (red) and observed (green) mean daily flow [m^3s^{-1}] of the best performing parameter set during calibration of Huai He Changtaigua, 1979 to 1991

There is generally good calibration over the full dataset with some underestimation of low-flow conditions with a resulting lower modified E result (0.86). Only the very highest flood flows are underestimated.

5.12.4 Zagunao He (Zagunao) Catchment Calibration

The Zaugunao He Zagunao sub-catchment is a 2520km^2 catchment in the western third of the YDB ranging in elevation from 1854m to 5357m. Again, following the methodology of calibration of the Bailu He, Laoguan He and Huai He sub-catchments, the results for the best performing parameter set for one year (1981), three years (1981 to 1983) and the full length of the calibration data available (1981 to 2000 with 1991, 1992 and 1994 missing) are given in tables 5.18, 5.19 and 5.20. The associated flow duration curves and full calibration daily flow are shown in figures 5.31, 5.32, 5.33 and 5.34.

Calibration	**C_{max}**	**c_r**	**c_l**	**c_{rb}**	**c_{lb}**	**R_r**	**R_l**	**k_d**
Zagunao He, Zagunao 1981	129.22	0.306	0.132	0.147	0.003	0.381	0.100	$2.23*10^{-6}$
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
-2.29	0.98	0.93	0.99	0.76	0.89	0.75	0.99	0.77

Table 5.18: Best performing parameter set and efficiency criteria results from calibration of the Zagunao He (Zagunao) catchment, 1981 (mean EC = 0.88)

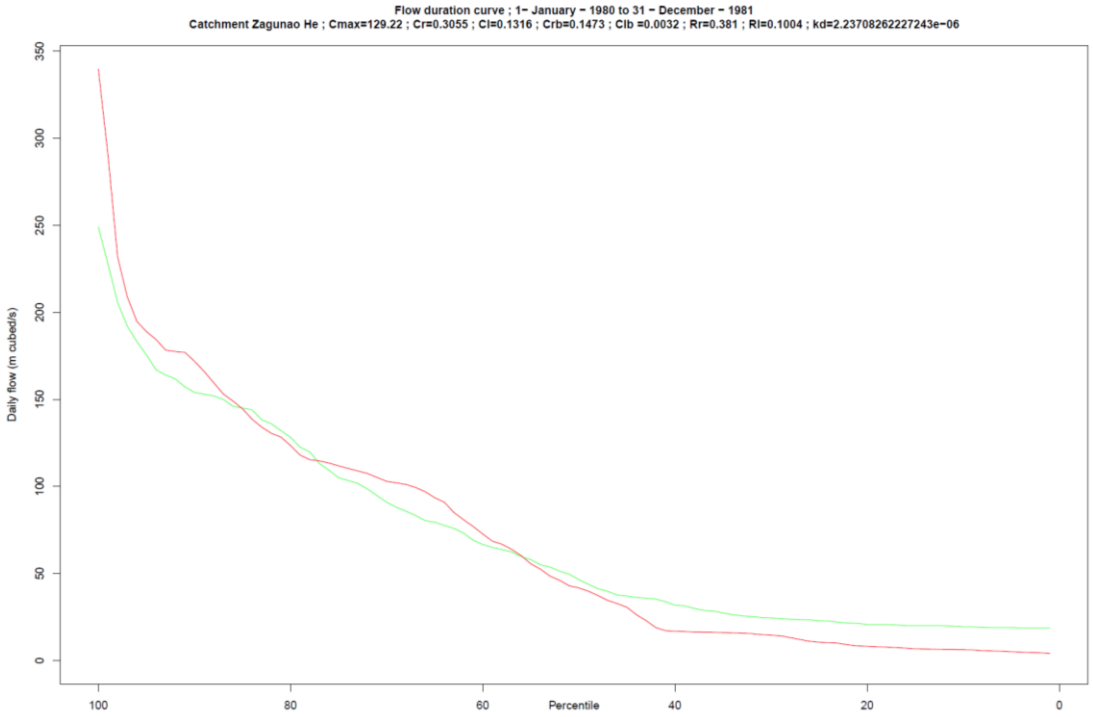


Figure 5.31: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Zagunao He Zagunao, 1981

Calibration	**C_{max}**	**c_r**	**c_l**	**c_{rb}**	**c_{lb}**	**R_r**	**R_l**	**k_d**
Zagunao He, Zagunao 1981-1983	227.523	1.559	1.062	0.005	0.088	0.443	0.178	$1.33*10^{-6}$
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
-5.92	0.99	0.98	0.99	0.84	0.93	0.81	0.95	0.81

Table 5.19: Best performing parameter set and efficiency criteria results from calibration of the Zagunao He (Zagunao) catchment, 1981 to 1983 (mean EC = 0.91)

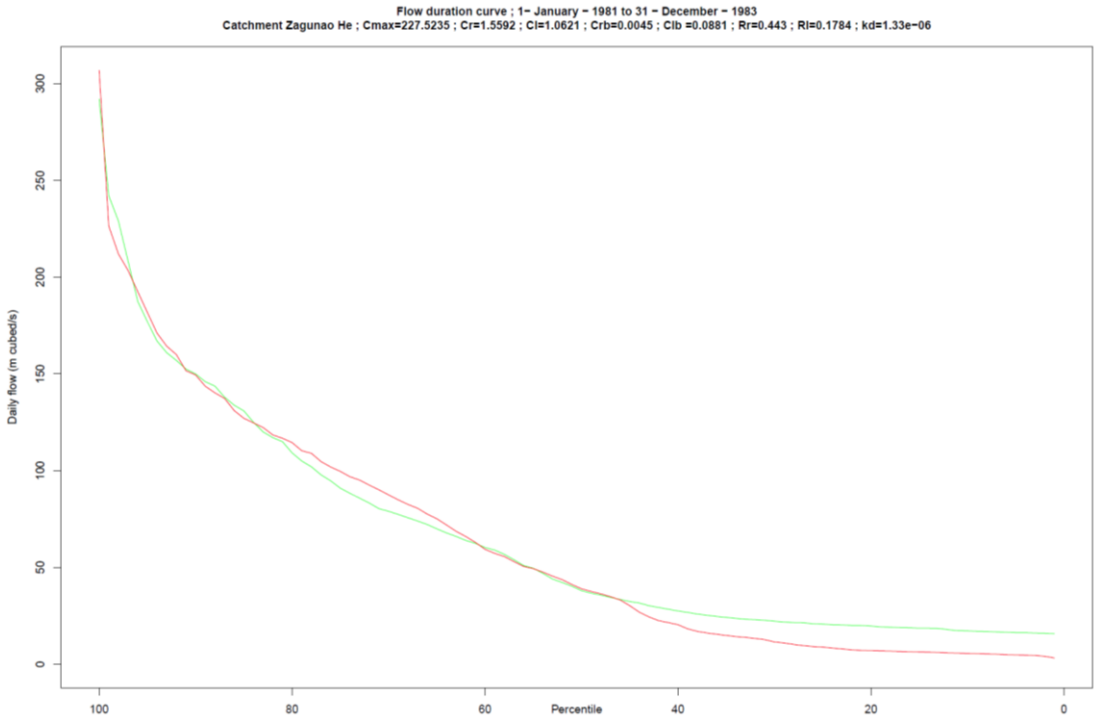


Figure 5.32: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Zagunao He Zagunao, 1981 to 1983

Calibration	**C_{max}**	**c_r**	**c_l**	**c_{rb}**	**c_{lb}**	**R_r**	**R_l**	**k_d**
Zagunao He, Zagunao 1981-2000	155.493	1.734	1.496	0.007	0.192	0.710	0.007	$5.95*10^{-6}$
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
-11.16	1.00	0.96	0.99	0.78	0.90	0.72	0.94	0.72

Table 5.20: Best performing parameter set and efficiency criteria results from calibration of the Zagunao He (Zagunao) catchment, 1981 to 2000 (mean EC = 0.88)

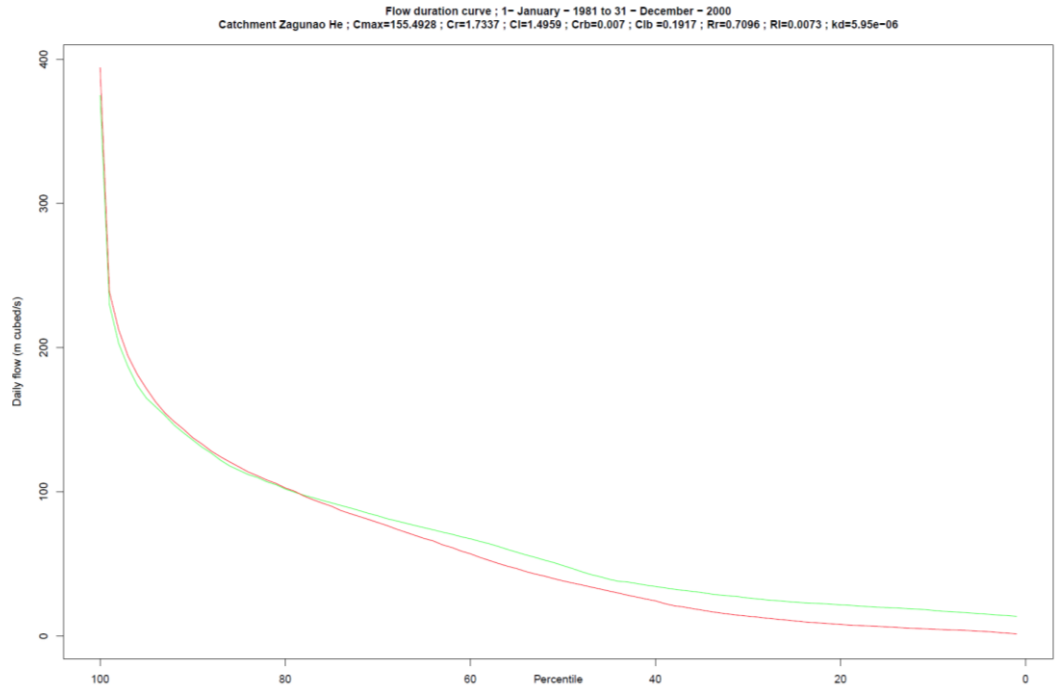


Figure 5.33: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the best performing parameter set during calibration of Zagunao He Zagunao, 1981 to 2000

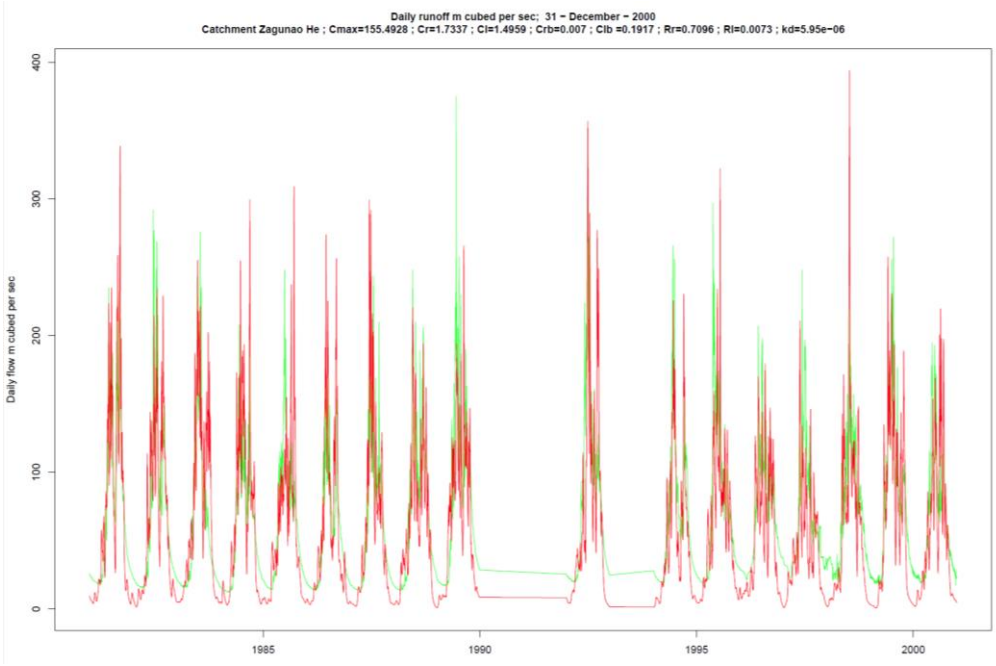


Figure 5.34: Modelled (red) and observed (green) mean daily flow [m^3s^{-1}] of the best performing parameter set during calibration of Zagunao He Zagunao, 1981 to 2000

There is a good fit of data between Q0 and Q25 but flows below Q25 are poorly represented by the model with resulting lower modified E, E_{rel} and E_{BoxCox} results (0.78, 0.72 and 0.72 respectively). It is unclear what the main reasons for this are, but suggestions include poorly represented input data for that catchment, inadequate representation of ground conditions for that catchment in the model or possibly requires parameters outside of the limits set.

5.12.5 Round-robin calibration of all catchments

The results obtained from sections 5.12.1 to 5.12.5 yield four optimised parameter sets from each of the four catchments calibrated, as shown in table 5.21. These parameter sets were used in a round-robin fashion to extract the modelling results for the six sub-catchments (i.e. Bailu He, Huai He, Lagunao He and Zagunao He plus the larger catchments Yalong Jiang Luning and Tongtian He Zhimenda) with view to finding a parameter set which gives an overall best model fit across the six sub-catchments.

Parameter Set Name	C_{max}	c_r	c_l	c_{rb}	c_{lb}	R_r	R_l	k_d
Bailu	120.577	0.256	0.883	0.066	0.039	0.448	0.150	$1.73*10^{-7}$
Laoguan	125.963	0.135	0.546	0.056	0.050	0.088	0.013	$4.53*10^{-7}$
Huai	42.499	0.905	0.312	0.031	0.051	0.005	0.079	$8.06*10^{-6}$
Zagunao	155.493	1.734	1.496	0.007	0.192	0.710	0.007	$5.95*10^{-6}$

Table 5.21: Best performing parameter sets following calibration of the four smaller sub-catchments

Initially, the four parameter sets were tested against the Yalong Jiang Luning sub-catchment from January 1979 to 1986 (eight years). With a catchment area of 107,882km^2, it is approximately 1.3 times the size of Scotland with elevations ranging from 1424m to 5814m. The results are shown in table 5.22:

Parameter Set Name	Vol Error %	r^2	E	d	Mod E	Mod d	E_{rel}	d_{rel}	E_{BoxCox}	Mean
Bailu	-48.9	0.94	0.64	0.92	0.30	0.69	0.21	0.82	-1.54	0.37
Laoguan	-31.5	0.97	0.85	0.96	0.55	0.80	0.44	0.86	-0.20	0.65
Huai	-29.3	0.98	0.86	0.96	0.58	0.80	0.58	0.89	0.25	0.74
Zagunao	85.3	0.98	-1.67	0.77	-0.32	0.59	0.24	0.94	-1.67	0.22

Table 5.22: Efficiency criteria performance of the four parameter sets during modelling of the Yalong Jiang Luning sub-catchment (107,882km^2), 1979 to 1986

Only the Laoguan and Huai parameter set calibrations give behavioural results although the low flow conditions are poorly modelled. The Zagunao parameter set does not calibrate and therefore was rejected from the round-robin. The best performing parameter set (Huai) FDC is shown in figure 5.35.

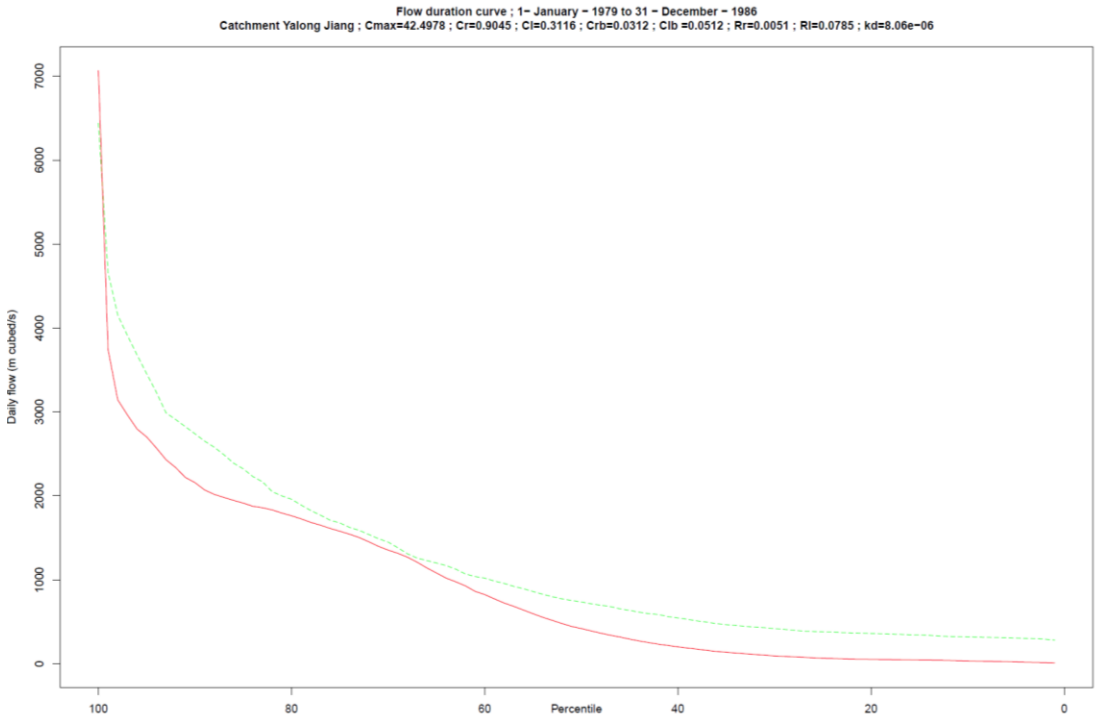


Figure 5.35: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the Huai parameter set used to model the Yalong Jiang Luning catchment, 1979 to 1986

The three remaining parameter sets were then tested against the remaining sub-catchments, including Tongtian He Zhimenda, a high elevation sub-catchment ranging from 3707m to 6452m with an area of 134,399km^2, and the results given in table 5.23.

The best performing parameter set was judged to be the one with the best combination of the overall mean of mean efficiency criteria and a mean of mean efficiency criteria weighted by area. A weighted mean score was calculated by multiplying the efficiency criteria for each catchment by the catchment area, summing these values for each parameter set and dividing through by the total area of all catchments (254,697km^2), as given in table 5.24.

Paramater Set Name	Vol Error %	r^2	E	d	Mod E	Mod d	E_{rel}	d_{rel}	E_{BoxCox}	Mean
Sub-catchment Tongtian He Zhimenda 137,704km^2 1979 to 1986										
Bailu	-37.5	0.97	0.85	0.96	0.57	0.79	0.64	0.90	0.31	0.75
Laoguan	22.3	0.99	0.81	0.97	0.64	0.85	0.88	0.98	0.87	0.87
Huai	5.9	0.98	0.97	0.99	0.83	0.92	0.90	0.98	0.91	0.94
Sub-catchment Bailu He Baiqueyua 290km^2 1979 to 1996										
Bailu	-14.5	0.95	0.90	0.97	0.88	0.94	0.99	1.00	0.98	0.95
Laoguan	-74.1	0.95	0.54	0.76	0.38	0.67	0.93	0.97	0.35	0.69
Huai	-51.2	0.91	0.80	0.92	0.57	0.79	0.93	0.97	0.48	0.80
Sub-catchment Laoguan He Xixia 3,216km^2 1979 to 1996										
Bailu	25.4	0.97	0.48	0.93	0.64	0.84	0.97	1.00	0.94	0.85
Laoguan	3.1	0.99	0.98	1.00	0.90	0.95	0.97	0.99	0.98	0.97
Huai	29.3	0.96	0.16	0.90	0.43	0.77	0.94	0.99	0.83	0.75
Sub-catchment Huai He Changtaigua 3,025km^2 1979 to 1991										
Bailu	25.0	0.97	0.96	0.99	0.75	0.88	-1.24	0.47	0.89	0.58
Laoguan	5.7	0.85	0.73	0.89	0.57	0.77	-1.26	0.05	0.82	0.43
Huai	4.6	0.97	0.97	0.99	0.86	0.94	0.97	0.99	0.92	0.95
Sub-catchment Zagunao He Zagunao 2,400km^2 1981 to 2000										
Bailu	-75.0	0.63	-0.29	0.74	-0.23	0.51	-0.37	0.72	-4.02	-0.29
Laoguan	-79.5	0.67	-0.40	0.68	-0.28	0.47	-0.38	0.68	-3.61	-0.27
Huai	-77.1	0.62	-0.36	0.72	-0.27	0.50	-0.43	0.71	-4.33	-0.36

Table 5.23: Round-robin efficiency criteria performance of three parameter sets (Bailu, Laoguan and Huai) during modelling of five sub-catchments

Parameter Set	Mean efficiency criteria	Weighted mean efficiency criteria
Bailu	0.54	0.58
Laoguan	0.56	0.76
Huai	0.64	0.84

Table 5.24: Mean and weighed mean of mean efficiency criteria

Table 5.24 clearly shows how the Huai parameter set performs best overall across the six sub catchments but performs poorly when modelling the non-behavioural Zagunao He sub-catchment. FDCs of the Tongtian He, Bailu He Laoguan He and Zagunao He sub-catchments modelled with the Huai parameter set are shown in figures 5.36 to 5.39.

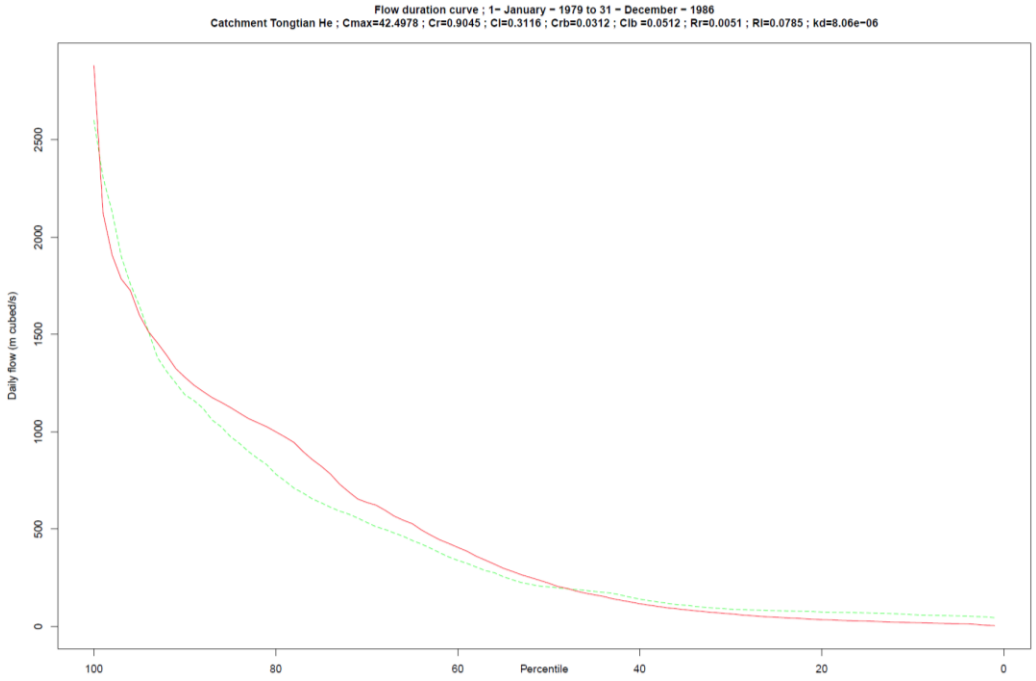


Figure 5.36: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the Huai parameter set used to model the Tongtian He Zhimenda catchment, 1979 to 1986

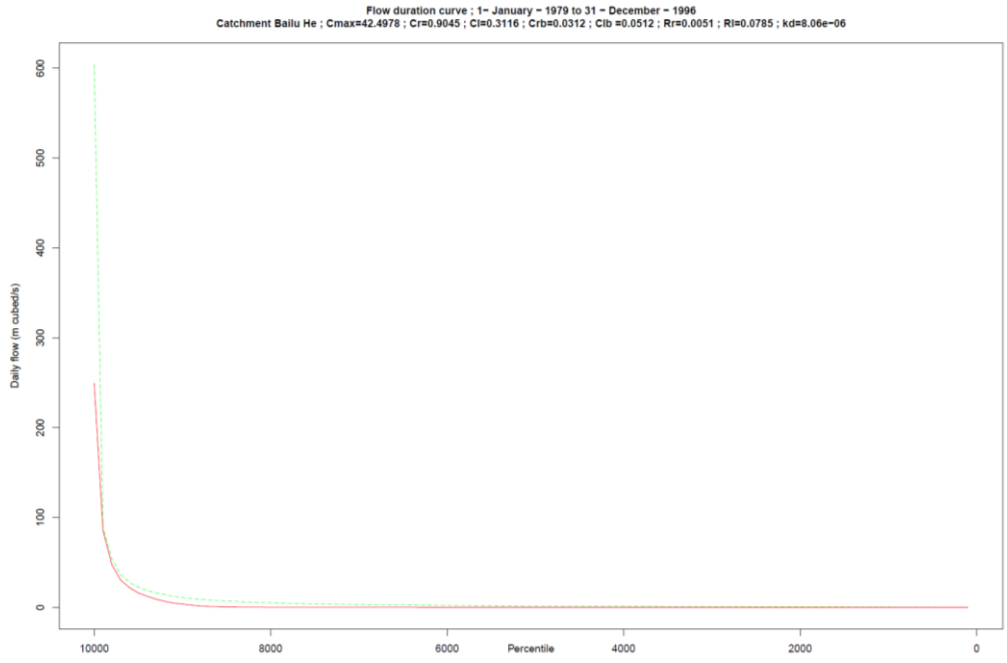


Figure 5.37: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the Huai parameter set used to model the Bailu He Baiqueyua catchment, 1979 to 1986

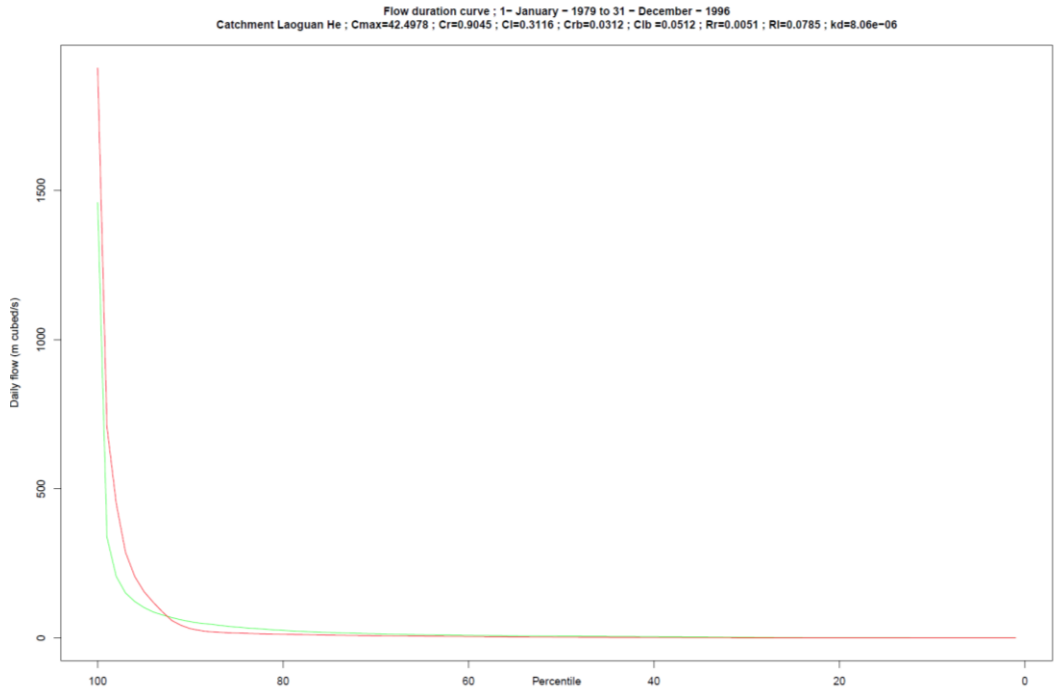


Figure 5.38: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the Huai parameter set used to model the Laoguan He Xixia catchment, 1979 to 1986

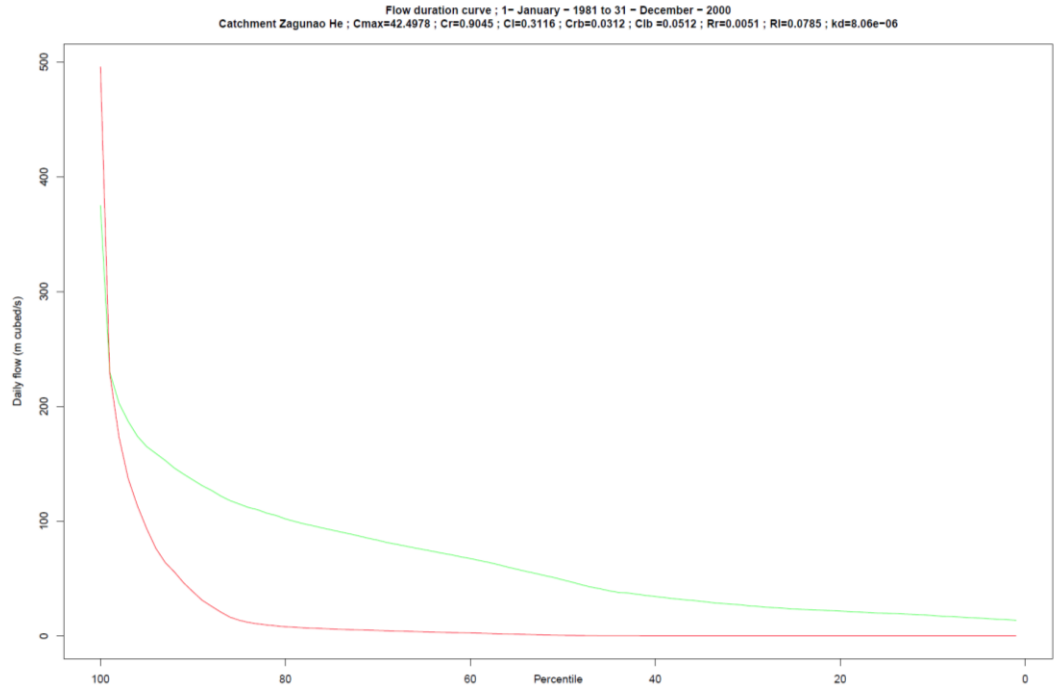


Figure 5.39: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] of the Huai parameter set used to model the Zagunao He Zagunao catchment, 1981 to 2000

5.12.6 Validation of the Huai He parameter set

To validate the suitability of the Huai He parameter set to model the whole of the YDB, the parameter set was used to model the Yangtze River Datong gauging station catchment. The Yangtze River Datong station catchment is 1,679,569km^2, ranging in elevation from 6452m to near sea level. Daily data for this catchment is only available for 2004 and the modelling time was approximately 14 days for one year. The resulting efficiency criteria results are shown in table 5.25 and FDC and mean daily flow plot shown in figures 5.40 and 5.41.

Validation	C_{max}	c_r	c_l	c_{rb}	c_{lb}	R_r	R_l	k_d
Yangtze River, Datong 2004 Model mean flow = 24,684m^{3}s^{-1}	42.499	0.905	0.312	0.031	0.051	0.005	0.079	8.06*10^{-6}
Vol Error %	**r^2**	**E**	**d**	**Mod E**	**Mod d**	**E_{rel}**	**d_{rel}**	**E_{BoxCox}**
-5.53	0.98	0.96	0.99	0.84	0.92	0.97	0.99	0.97

Table 5.25: Validation of the Huai parameter set through modelling the Yangtze River Datong catchment, 2004 (mean EC = 0.95)

Mean efficiency criteria of 0.95 indicates a good model fit with the Huai parameter set.

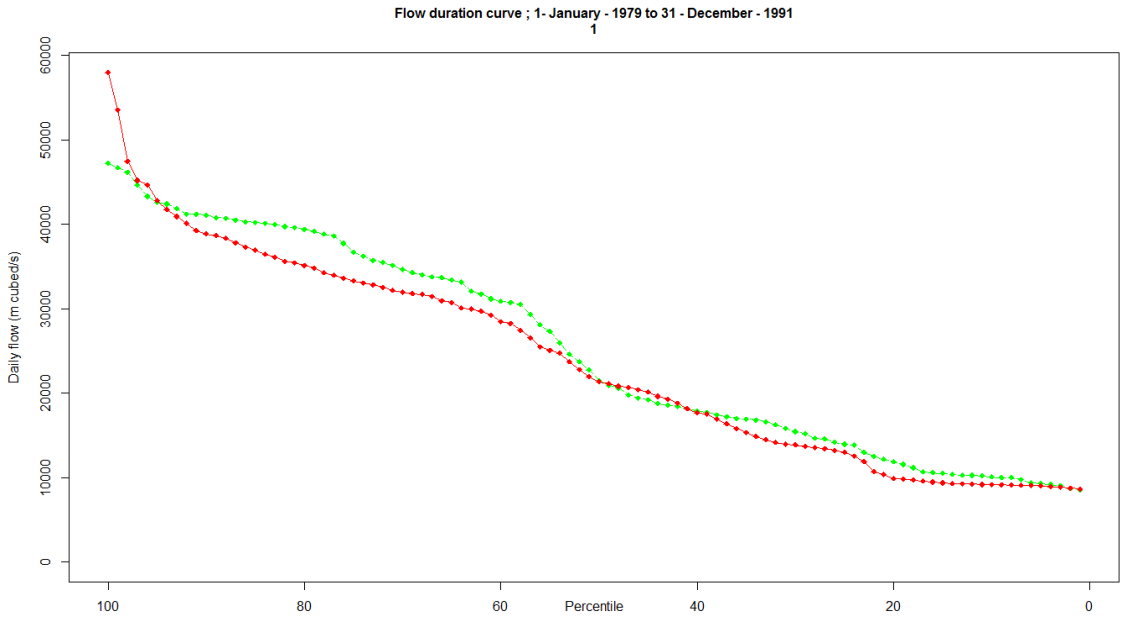


Figure 5.40: Modelled (red) and observed (green) flow duration curve [m^3s^{-1}] using the Huai parameter set to model the Yangtze River Datong catchment, 2004

Figure 5.41: Modelled (red) and observed (green) mean daily flow [m^3s^{-1}] using the Huai parameter set used to model the Yangtze River Datong catchment, 2004

5.13 Extracting mean daily flow data of the YDB from 1979 to 2007

Although the calibration data for the Yangtze River Datong catchment is limited (2004 only), the mean efficiency criteria results established through validating the Huai dataset indicate a good model fit across the catchment overall and across smaller sub-catchments. However, there are exceptions such as the Zagunao He Zagunao sub-catchment. The Huai parameter set was then employed to model the whole of the YDB from 1979 to 2007. Ideally the YDB would be modelled in one continuous time-series so the data from the end of one year would feed into the next, but with the time taken for one year (approximately 14 days) it would take approximately 1.1 years to complete. This is longer than the time available and it is unlikely that a single computer would be stable enough for this length of time.

Therefore, it was decided to run the 29 years in parallel on 29 computer cores using the Eddie Linux Compute Cluster Mark 3 i.e. each year was ran in isolation and as 30+ cores can be ran at the same time by a single user, the total modelling time was cut from 1.1 years to just over 14 days. The output from this was 10,592 mean daily flow raster .tif files, each approximately 48MB in size.

The modelled mean daily flow [m^3s^{-1}] at the Yangtze River Datong station is shown in figure 5.42, with flow varying from 4,106m^3s^{-1} to 80,423m^3s^{-1}.

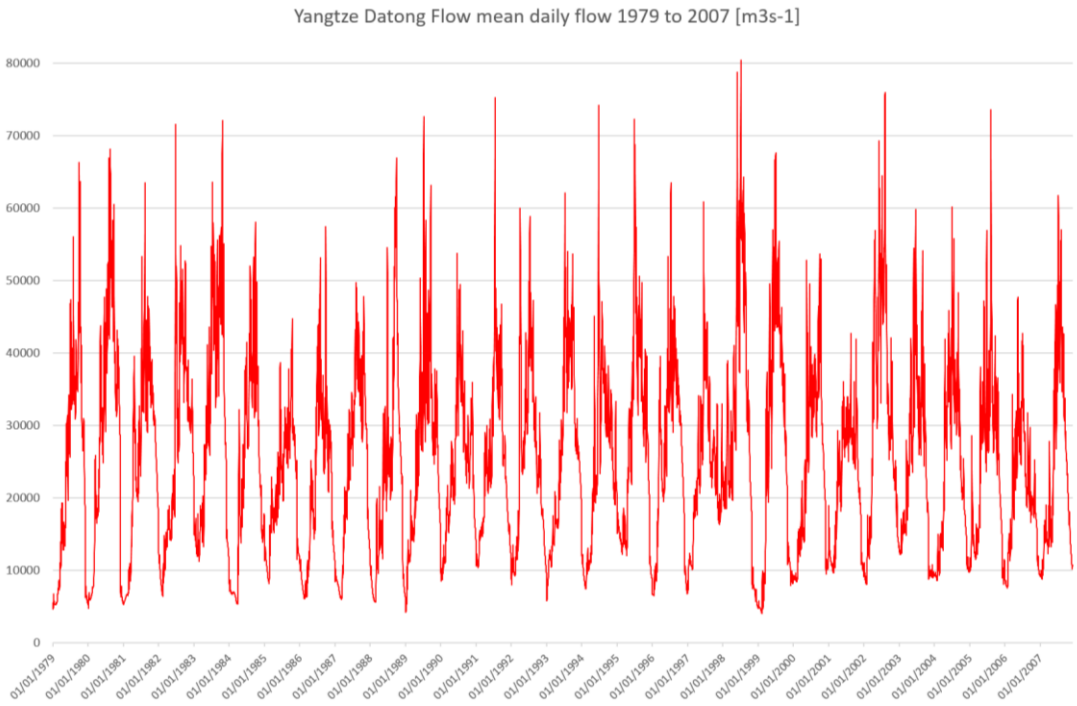


Figure 5.42: Modelled mean daily flow at the Yangtze River Datong gauging station 1979 to 2007 [m^3s^{-1}]

5.14 Extracting FDC data for the whole of the YDB

For the hydropower search algorithm it is necessary to create rasters representing flow quantile information of the YDB in 5% intervals i.e. rasters representing Q_0, Q_5, Q_{10}, Q_{15}…..Q_{85}, Q_{90}, Q_{95} and Q_{100}, where Q_0 represents the peak/flood flows and Q_{100} the very low flows. This can be achieved by stacking the 10,592-mean daily flow raster files and extracting quantile information of each cell column. A raster stack is a collection of raster objects with the same spatial extent and resolution created using the *Raster* package *stack* tool. Quantile data is then extracted using the *Raster* package *quantile* tool, setting a probability parameter between 0 and 1 (where 1 represents Q_0 and 0 represents Q_{100}).

However, stacking 10,592 48MB files requires over 500GB of memory which very few systems have. Therefore, each mean daily flow raster image (each with a resolution of 7155*2710 cells) was split into 150 tiles (each with a resolution of 477*271 cells) with view

to creating 150 tile columns, easing the memory burden to under 4GB per tile column. The quantile data of each column was then merged together to create 21 quantile raster images and, to minimise the storage space, data over land cells was set to NA. Of the 19,390,050 cells of each raster, 1,411,299 (7.3%) are classed as river cells.

5.15 Parameter uncertainty

Model calibrations and subsequent predictions are subject to uncertainty as no hydrological model is a true reflection of the processes involved (Bevan, 2012). There are uncertainties in model structures, parameter estimates, initial and boundary conditions and input and observation data. Different calibration datasets and performance measures result in differing optimum parameter sets. Model parameter uncertainty can be characterised using the General Likelihood Uncertainty Estimation (GLUE) method (Beven, 1992) to estimate the degree of belief associated with differing models and parameter sets, recognising the concept of equifinality. This can be seen in dotty plots of a single efficiency criteria (Nash Sutcliffe efficiency used in this example) plotted against parameters values (figure 5.43), where each dot represents one model run. Both good and poor fitting models are found across the range of each parameter, and therefore the model fit is not a function of individual parameters but the interaction between components of the parameter set.

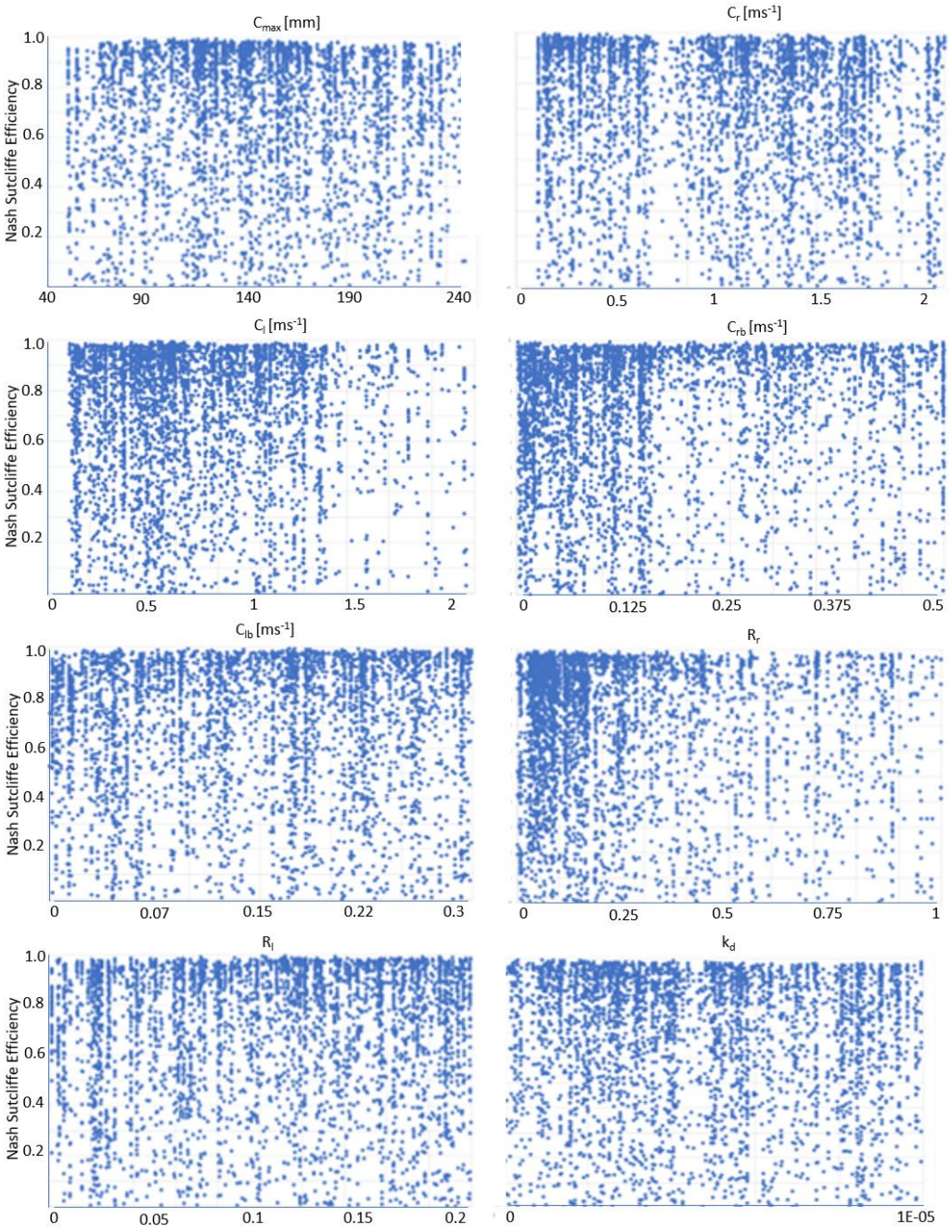


Figure 5.43: Dotty plots of each model parameter vs Nash Sutcliffe efficiency showing concept of equifinality

The GLUE methodology was applied to the calibration results of the Bailu He Baiqueyua catchment. Models with a mean efficiency criteria of ≤ 0.5 were rejected as non-behavioural. Predictions of each simulation were weighted by the efficiency criteria associated with each simulation, and the cumulative likelihood weighted distribution of predictions used to estimate 5% and 95% quantiles for the predictions at any timestep. This shows the impact different reasonable parameters will have on the predicted hydrograph (Duncan, 2014) (see figure 5.44). This shows that uncertainty is greatest during peak flows suggesting issues with either the input data or the model structure, but baseflow is well captured.

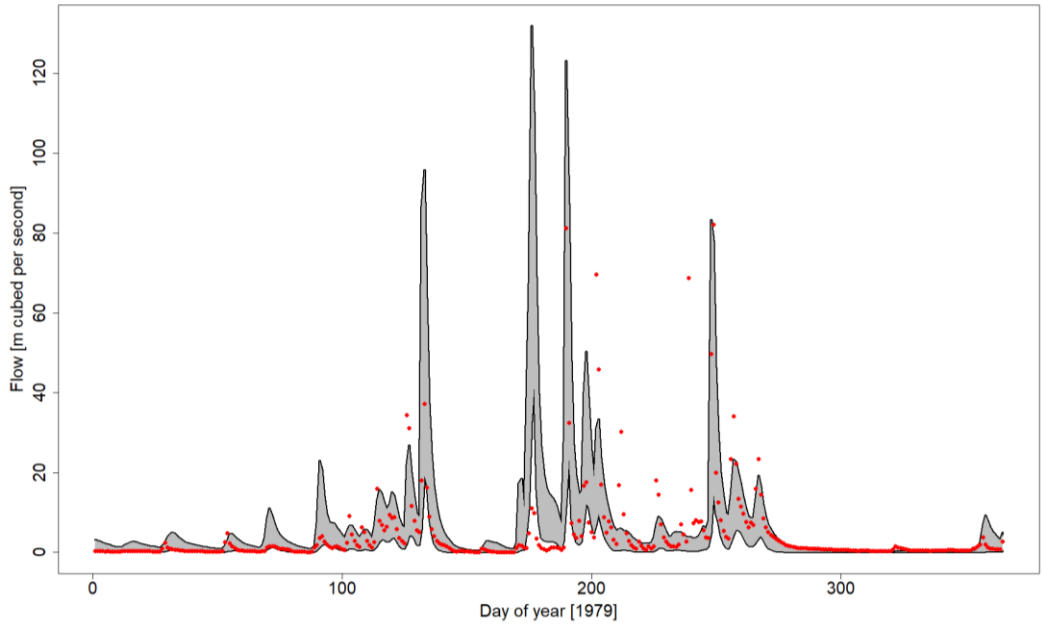


Figure 5.44: Parameter uncertainty using the GLUE methodology for the Bailu He Baiqueyua catchment during 1979

5.16 Chapter conclusions

- Due to limitations in space and time of river flow measurements, hydrological models are required to extrapolate other available data.
- Hydrological models are classified based on their structure, and considered lumped or distributed, usually requiring calibration to establish optimal process parameters by comparing to an objective function.

- G2G (Grid to Grid) was selected as an appropriate basis for a hydrological model as it can simulate flows over large domains at high-resolution and able to be integrated with GIS datasets, but only available for commercial use.
- The equations that form the basis of the G2G model were published and therefore could be included in a new hydrological model incorporating the selected DEM and meteorological datasets.
- G2G has a run-off production model based on water inputs and outputs and a routing model based on simplified St.Venant equations and has eight parameters requiring calibration.
- A graphical interface distributed hydrological model at 15 arc-seconds resolution was developed in the R programming language, incorporating the G2G equations and the SNOW-17 model, with outputs produced as daily hydrographs and flow duration curves as PDF files, and process data saved to an external database.
- Simulating one year of the YDB takes approximately 2 weeks which could result in difficulties simulating a 30-year record.
- To calibrate the model, the model results were compared to observations across 8 efficiency criteria in conjunction with the volume error, as multi-objective efficiency criteria deemed to be stronger than a single objective efficiency measure.
- The parameter ranges were constrained by results of previous uses of the model, with view to reducing the number of model runs to achieve a good calibration result.
- A genetics based evolutionary algorithm (DEoptim) was selected to optimise the process parameters due to it consistently minimalising efficiency criteria compared to other packages and its ability to run in parallel.
- A round-robin calibration strategy was developed initially calibrating sub-catchments of the YDB, partly due to the model runtimes of the YDB and partly due to limited observational data.
- Validation of the final selected parameter set was achieved by comparing observation results compared to model results on a 1,679,569km^2 sub catchment of the YDB across 2004. This resulted in a mean efficiency criteria result of 0.97, suggesting the hydrological model is sufficient for extraction of flow values for use in a hydropower search engine.

- Due to time constraints running the whole YDB catchment for 30 years, each year was run in parallel. Although not ideal, this was deemed an acceptable compromise with good efficiency criteria results for 2004.
- Daily flow rasters were extracted and used to create FDC data for each river point across the YDB to produce 21 quantile rasters at 5% intervals, as input to the hydropower search algorithm.
- Employing the GLUE methodology suggested that parameter uncertainty was greatest during peak floods with baseflow well captured. As hydropower schemes are less likely to be designed to operate with peak flow volumes, overall the hydrological model performed well for the purpose.

6 Development of a hydropower search algorithm

6.1 Introduction

This chapter discusses the development of a hydropower search algorithm based on the RETScreen Small Hydro Project Model as a source of power production and costing equations, utilising FDC, terrain and other datasets derived in the previous chapters. The aim of the algorithm is to search the YDB river network to determine the power resource available and estimate costs to determine those schemes with a sufficient internal rate of return (IRR) to be viable. RETScreen's Small Hydro Project Model is primarily aimed at assessing ROR projects and hence impoundment type projects will not be assessed, although there are many such examples within the YDB. Coding of the algorithm was developed entirely in the R programming language.

6.2 An introduction to RETScreen

RETScreen (2004) is a clean energy management software suite designed for energy efficiency, renewable energy and cogeneration project feasibility analysis as well as ongoing energy analysis. The software was published by Natural Resources Canada, covering most renewable energy system types. RETScreen's main aims are to empower energy professionals and decision makers to rapidly identify, assess and optimise the financial viability of clean energy projects, as well as verify the performance of facilities (Natural Resources Canada, 2016). The latest version published in 2016, RETScreen Expert, is now a commercial software suite and designed to assess a single scheme.

Detailed descriptions of the algorithms of the RETScreen model were published in the document *RETScreen International Clean Energy Project Analysis Engineering and Cases Textbook* (RETScreen, 2005). Power production and costing equations are found in an appendix to the Small Hydro chapter which were utilised to generate a new hydropower search algorithm designed to search the whole YDB catchment.

The Small Hydro model is primarily designed to assess ROR projects up to 50MW i.e. those with no or limited pondage (or water storage). It is important to note that it is assumed that the RETScreen methodology is also applicable above 50MW, but it is quite possible that there are limitations of the model when used for schemes beyond 50MW. Projects are labelled as ROR differently across the world, ranging from schemes with no storage to those with a large reservoir. Projects are sometimes labelled ROR to placate public perception. The Belo Monte Dam in Brazil is an 11,233MW scheme classed as ROR yet with a reservoir storage surface area of 440km^2.

6.3 Validation of the RETScreen Methodology

RETScreen's Small Hydro Project Model was validated by hydropower experts, cost engineering experts, greenhouse gas modelling specialists, financial analysis professionals, weather scientists and hydrologists. Examples of the validation include:

1. Comparison of RETScreen's algorithm to determine turbine efficiencies to the manufacturer guaranteed turbine efficiency for the Brown Lake Hydro Project in British Columbia, Canada, resulting in a good approximation of the as-designed turbine efficiencies (see figure 6.1).
2. Comparison of the plant capacity and annual renewable energy delivered between the RETScreen Small Hydro Project Model and an alternative software (Hydra) presented in a report for the International Energy Agency - Implementing Agreement for Hydropower Technologies and Programmes entitled "Assessment Methods for Small-hydro Projects" (Wilson, 2000), which found little difference between the estimates of both models.
3. Comparison of RETScreen Small Hydro Project Model initial costs against detailed actual initial costs for the Rose Blanche Hydroelectric Development, Newfoundland, Canada. RETScreen overestimated the costs by 14%. This is unsurprising as each hydropower project is bespoke and costs will depend on specific local conditions, however at the pre-feasibility stage this discrepancy is considered acceptable.

Of course, data inputs to the RETScreen Model are also subject to error. It is important to recognise that different sized hydroelectric schemes differ significantly in approach and scale, impacting on the economics of a scheme.

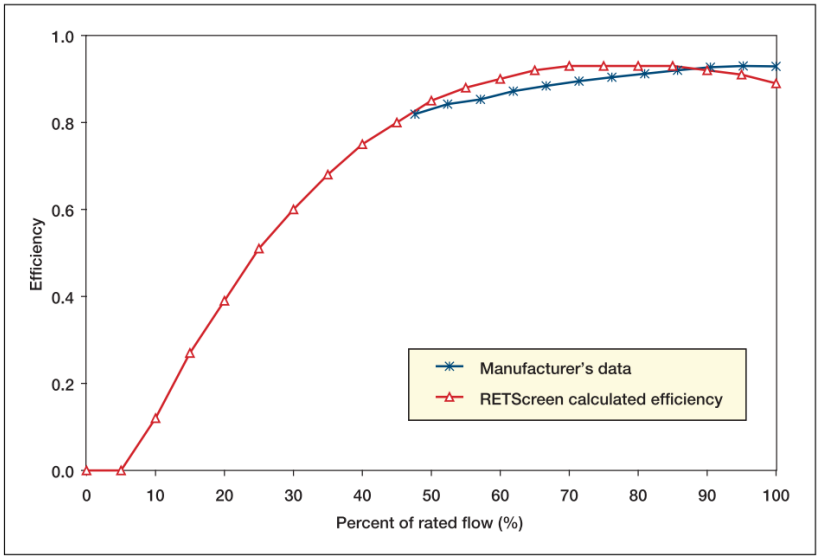


Figure 6.1: Comparison of RETScreen calculated hydro turbine efficiency against manufacturer's data (RETScreen, 2005)

6.4 Developing a catchment hydropower search algorithm

RETScreen software is used primarily to assess the pre-feasibility of one scheme, possibly with a few iterations of design flow or turbine type etc. However, it is not designed to search an entire catchment with many iterations of turbine type, intake and turbine location and design flow for each river point, which is the aim of this project. Therefore, a new model was developed within this project by the author, coded in the R programming language, to scan the river network of a catchment incorporating RETScreen's power production and costing formulae and FDC and terrain datasets developed earlier.

In developing a search algorithm, it is assumed that each site is a simple ROR hydropower scheme consisting of a weir (or dam), a penstock, turbine and generator contained in a powerhouse and a tail race, with energy transmitted to the local or centralised network via a powerline. Virtual access roads are constructed for each scheme linking the existing road network to both the intake and turbine. Larger schemes may have multiple penstock/turbine configurations.

Hydropower projects are site specific and therefore assumptions were necessary in developing the new model, including:

- Each scheme is assumed to be ROR, with no costings included for pondage (water storage) i.e. the weir is not designed to store water but to channel the water into a penstock.
- The RETScreen methodology is acceptable in assessing schemes greater than 50MW.
- Residual flow is set to Q_{95} (i.e. flow at Q_{95} or below is classed as 'hands off'). Note that some countries have variable residual flow laws to better simulate natural river flow variation.
- River points assessed are at the centre of each 15-arcsecond grid square.
- An access road is created for each site linking to the existing road network. Roads are created to reach the intake and turbine.
- Aqueducts and tunnels are not considered.
- Each scheme has a new transmission line.
- Electrical losses are set at 2% for the transmission line and 2% for the generator.
- A single penstock material (glass reinforced plastic (GRP)) is considered.
- A scheme can have up to a maximum of 100 penstock/turbine configurations.
- Electricity prices are fixed and all the energy produced can be used/sold.
- Construction is assumed to be complete by the start of the project (i.e. year 0).

The last point is important. Normally finance is required during the construction phase which could last several years. Assuming projects begin at year 0 removes complicated financial payments, debt and interest accumulation during construction, and therefore the hydropower search algorithm is solely comparing costs of schemes as if they were built instantaneously. Large hydropower schemes would require significantly more construction time than small schemes, and hence the actual costs of an individual scheme may be significantly more than detailed here. Hence, construction time should be considered during a more detailed investigation of a potential scheme.

6.5 A description of the hydropower search algorithm

Similar in style to the hydrological model developed in chapter 5, the hydropower search algorithm is composed of several R functions which are called when appropriate (see figure 6.2) and are described in the following sections.

6.5.1 Installing packages and loading libraries

As described in 5.4.1, several R packages need to be installed for the model to run, and available as .tar compressed files available via CRAN. The *Load libraries* function loads the requisite packages which include:

- *raster* - (see 5.6.1)
- *maptools* - (see 5.6.1)
- *sp* - (see 5.6.1)
- *rgeos* - (see 5.6.1)
- *rgdal* - (see 5.6.1)
- *proj4* (Urbanek, 2012) – A simple interface to latitude/longitude projection and datum transformation of the PROJ.4 cartographic projections library. It allows transformation of geographic coordinates from one projection and/or datum to another
- *gdistance* (van Etten, 2017) - Calculate distances and routes on geographic grids
- *rootSolve* (Soetaert, 2016) - Includes routines that: (1) generate gradient and Jacobian matrices (full and banded), (2) find roots of non-linear equations by the 'Newton-Raphson' method, (3) estimate steady-state conditions of a system of (differential) equations in full, banded or sparse form, using the 'Newton-Raphson' method, or by dynamically running, (4) solve the steady-state conditions for uni-and multicomponent 1-D, 2-D, and 3-D partial differential equations, that have been converted to ordinary differential equations by numerical differencing (using the method-of-lines approach). Includes Fortran code
- *geosphere* (Hijmans et al., 2016b) - Spherical trigonometry for geographic applications. That is, compute distances and related measures for angular (longitude/latitude) locations
- *FNN* (Beygelzeimer et al., 2013) - Fast Nearest Neighbour Search Algorithms and Applications, cover-tree and kd-tree fast k-nearest neighbour search algorithms and related applications including KNN classification, regression and information measures are implemented
- *FinCal* (Fan, 2016) - Package for time value of money calculation, time-series analysis and computational finance

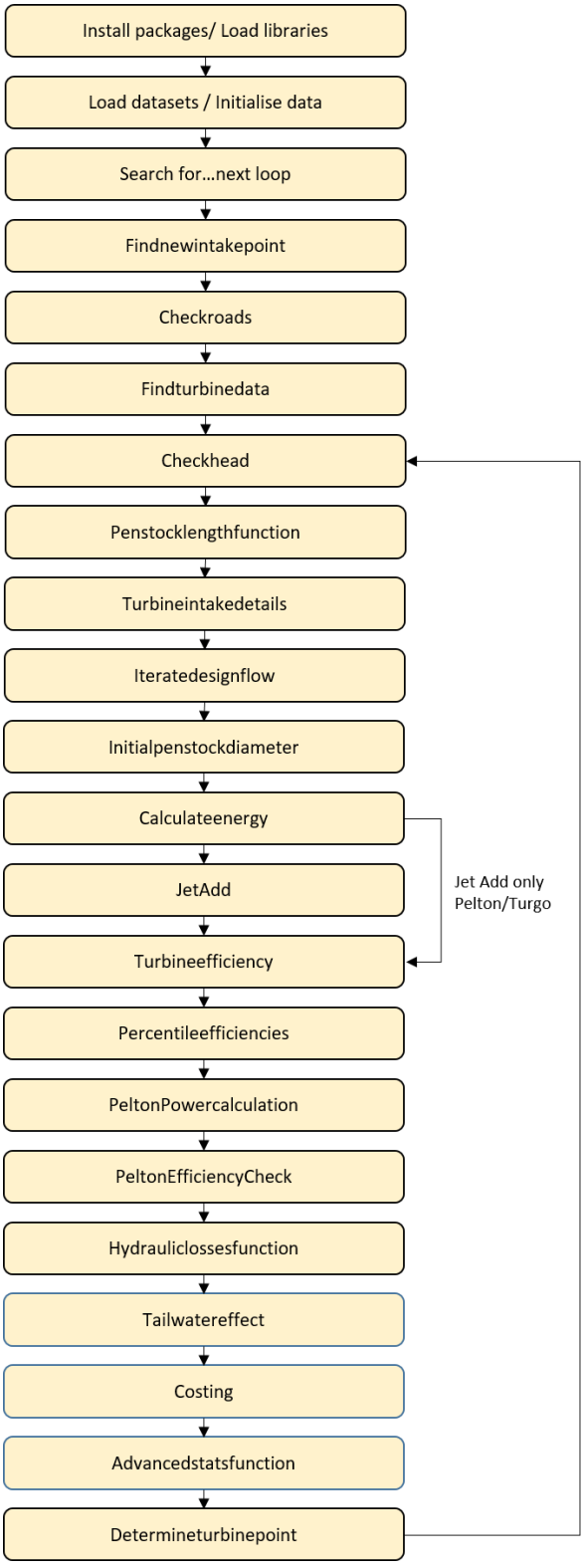


Figure 6.2: Flow diagram of the hydropower search algorithm functions

- *data.table* (Dowle et al., 2017) - Fast aggregation of large data (e.g. 100GB in RAM), fast ordered joins, fast add/modify/delete of columns by group using no copies at all, list columns, a fast friendly file reader and parallel file writer. Offers a natural and flexible syntax, for faster development
- *plyr* (Wickham, 2016) – Tools for splitting, applying and combining data.

6.5.2 Loading datasets and initialising data

Input datasets include the DEM, FDR and slope raster files that were developed in chapter 3, and the 21 FDC (Q_0 to Q_{100} in 5% intervals) raster files that were developed in chapter 5. Other necessary inputs include the Yangtze road network shapefile, Yangtze lakes and reservoirs shapefile and the 220kV+, 330kV+ and 500kV+ transmission network shapefiles, all developed in Chapter 3. The villages/towns/cities, towns/cities and cities only shapefiles developed in chapter 3 were loaded into the model and converted to a data frame table consisting of longitude and latitude data.

During development of the hydropower search algorithm it was found that the Yangtze road network shapefile was too dense, making any distance calculations too slow to be useful. Therefore, the Yangtze road network polygon shapefile was converted to a points shapefile by:

1. Converting the Yangtze road network polygon shapefile to a raster file at 15-arcsecond resolution using the 'Polygon to raster' tool within ArcGIS.
2. Converting the Yangtze Road network raster at 15-arcsecond resolution to a points shapefile using the 'Raster to point' tool within ArcGIS.

This new shapefile has a point in the centre of any grid square instead of polygon lines crossing the grid square (see figure 6.3), improving the speed when utilising the dataset. However, this is a compromise as the distance to roads from the river network now has a potential error of up to +/-7.5-arcseconds (approximately +/-225m).

A final river network raster to be analysed was derived by using the Q_{mean} data, which was filtered to remove river flow values < 10 ls^{-1}. The remaining river points were set to a value of 1 (see figure 6.4). Summing the river points remaining yielded 1,411,299 river points to be analysed across the YDB.

Figure 6.3 Example section of the original dense road polygon shapefile (blue line) and new lighter road points shapefile (red dots)

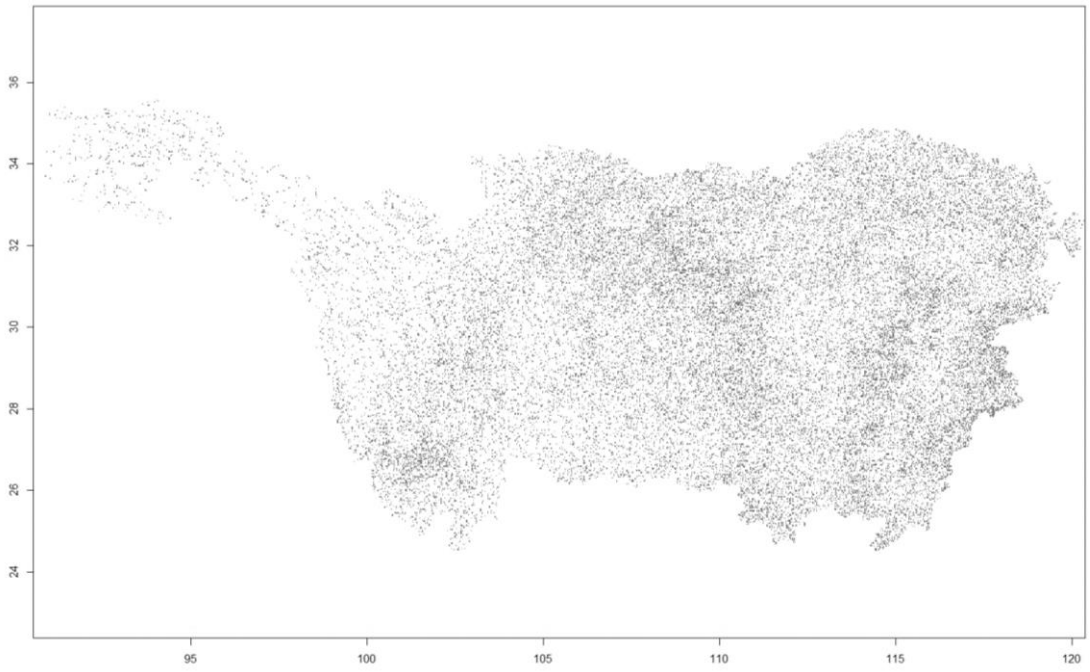


Figure 6.4: Final river flow network to be analysed (1,411,299 river points)

The flow available [m^3s^{-1}] to a hydropower scheme at each percentile required subtraction of the 'hands-off' flow as given in equation 6.1, where n represents the percentile flow:

$$\mathbf{Q_{n.available}} = \frac{\mathbf{Q_n - Q_{95}}}{\mathbf{1000}} \mathbf{[m^3s^{-1}]} \qquad (6.1)$$

Turbines are selected based on the available head (difference in elevation between intake and turbine) and the available flow rate, where impulse turbines are used for high head sites and reaction turbines for low head sites. Turbines considered in the RETScreen Small Hydro Model include the Pelton, Turgo and Cross-flow impulse turbines, and the Francis, Kaplan and Propeller reaction turbines. To select an appropriate turbine requires a turbine selection chart, a series of polygons depicting the flow/head range of each turbine. Due to the large range of heads and flows available to some of the turbines available, turbine selection charts are normally expressed as a logarithmic chart (see figure 6.5).

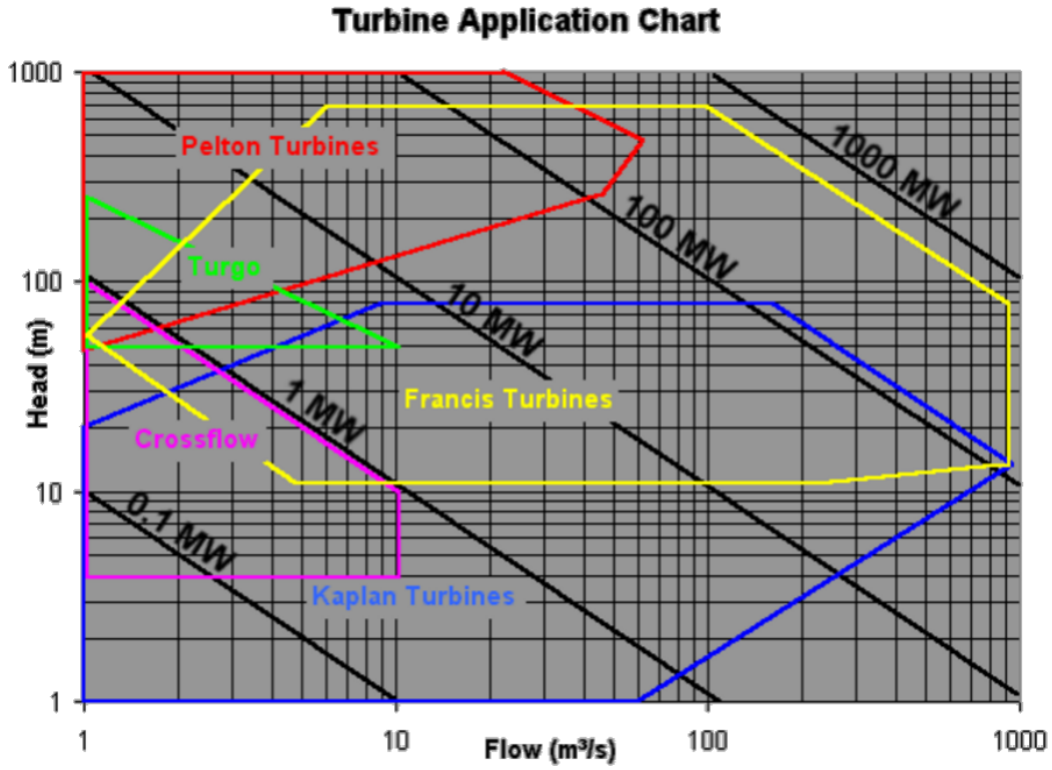


Figure 6.5: A logarithmic turbine application chart

To enable selection of appropriate turbines within the model, a polygon plot of each turbine was created with the vertices corresponding to those in the logarithmic plot in figure 6.5. This was then transformed into a Spatial Polygons Data Frame (using the *SP* package) - effectively

a turbine map that can be used to select appropriate turbines based on design flow and head (figure 6.6). The Francis reaction turbine (yellow polygon) has the largest range with others suitable for low flow and/or low/high head conditions. It is assumed that the Kaplan and Propeller turbines have the same flow/head range as each other. Table 6.1 shows the polygon colour for each of the possible turbines in the plot:

Colour of Polygon	Turbine
Yellow	Francis (Reaction) turbine
Red	Pelton (Impulse) turbine
Green	Turgo (Impulse) turbine
Blue	Kaplan and Propeller (Reaction) turbines
Magenta	Crossflow (Impulse) turbine

Table 6.1: Polygon colours of different available turbines in figure 6.7

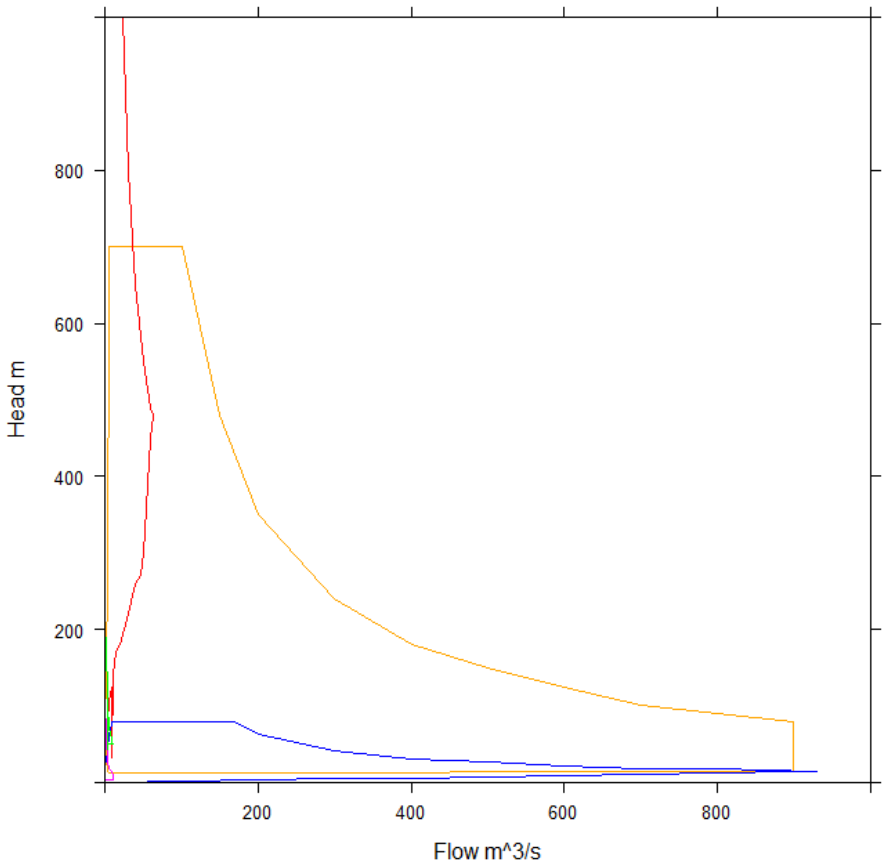


Figure 6.6: Transformation of the logarithmic turbine application chart to a non-logarithmic Spatial Polygons Data Frame (a turbine map)

To put energy generation in context, the model captures the number of homes that could be powered by each scheme by using the average energy use for a UK and Chinese home, which are set at 4,648kWhr and 1,349kWhr per annum respectively (World Energy Council, 2010). Hydropower schemes are categorised depending on their generation capacity, and the following table (6.2) gives the number of homes each category of hydropower scheme could power:

Classification of hydropower scheme	**Generation capacity range**	**UK homes powered (based on World Energy Council usage (2010))**	**Chinese homes powered (based on World Energy Council usage (2010))**
Large	>50MW	>95 thousand *	>325 thousand *
Small	≥ 1MW & ≤ 50 MW	1,884 to 94,234	6,493 to 324,684
Mini	> 100kW & < 1MW	188 to 1,884	649 to 6,493
Micro	≥5kW & ≤100kW	9 to 188	32 to 649
Pico	≥ 0.5 kW & < 5kW	1 to 9	3 to 32

* the upper limit of large hydro power schemes is currently 22GW which is the potential output of the Three Gorges Dam in China, currently the largest power station (of any form) in the world

Table 6.2: Categorisation of hydropower schemes and number of homes powered

The RETScreen costing formulae were developed in 2004 in Canadian dollars. Following the methodology of Duncan (2014), the costs were re-calculated to Chinese Yuan (CNY) in 2016 by multiplying by C_{con} which includes two correction factors; an exchange rate (C_{ex}) and the Chinese rate of inflation (C_{inf}). Costs are calculated in Canadian dollars, converted to Chinese Yuan at the 2004 exchange rate C_{ex} (1 Canadian dollar = 6.387 Chinese Yuan (UKForex, 2016)) and then inflated at the average inflation rate between 2004 and 2016, i_{inf}, (2.875% (World Bank, 2017)) over the 12 years between 2004 and 2016 using:

$$\mathbf{C_{inf} = (1 + i_{inf})^{12}} \qquad (6.2)$$

$$\mathbf{C_{con} = C_{inf} \times C_{ex} = 8.975} \qquad (6.3)$$

RETScreen assumes that hydropower schemes in areas of hard ground frost (due to latitude, elevation or during colder periods of the year) have additional costs due to the difficulty in working with frozen ground. Therefore a 'Number of Frost Days' raster (figure 6.7) was created, assuming a frost-day is one where the average temperature is below 0°C, using the following methodology:

1. All the temperature dataset files for 1979 (APHRODITE, 2016) were assembled at the native 0.25-degree grid resolution (i.e. 365 files, one for each day) into a raster stack.
2. The resolution of each was increased to that of the DEM (i.e. from 0.25 degrees to 15-arcsecond) using the disaggregate function.
3. The resolution of the DEM was decreased to that of the original temperature dataset (i.e. from 15-arcseconds to 0.25 degrees) using the aggregate function.
4. The aggregated DEM resolution was increased back to the original resolution (i.e. from 0.25 degrees back to 15-arcseconds). This is now a DEM at the original high-resolution but the data averaged to the lower resolution.
5. The difference in height was found between the original high-resolution DEM and the averaged DEM.
6. The calculated height difference found in step 5 was multiplied by the adiabatic lapse rate.
7. An adjusted high-resolution (15-arcsecond) temperature stack was re-calculated by adding the temperature data calculated in step 6 to each of the increased resolution temperature files found in step 2.
8. For each cell column within the adjusted temperature stack, the number of days that have an average temperature less than 0°C was found (see figure 6.8).

A frost-days 'F' factor is then calculated:

$$\mathbf{F} = \frac{\mathbf{110}}{(\mathbf{365}-\mathbf{f})^{\mathbf{0.9}}} \qquad (6.4)$$

where f is the number of frost days, and if f is equal to 365 then f is set to 364.9. Rearranging equation 6.4 gives an F factor value of 1 when the number of frost days, f, is equal to 180 days.

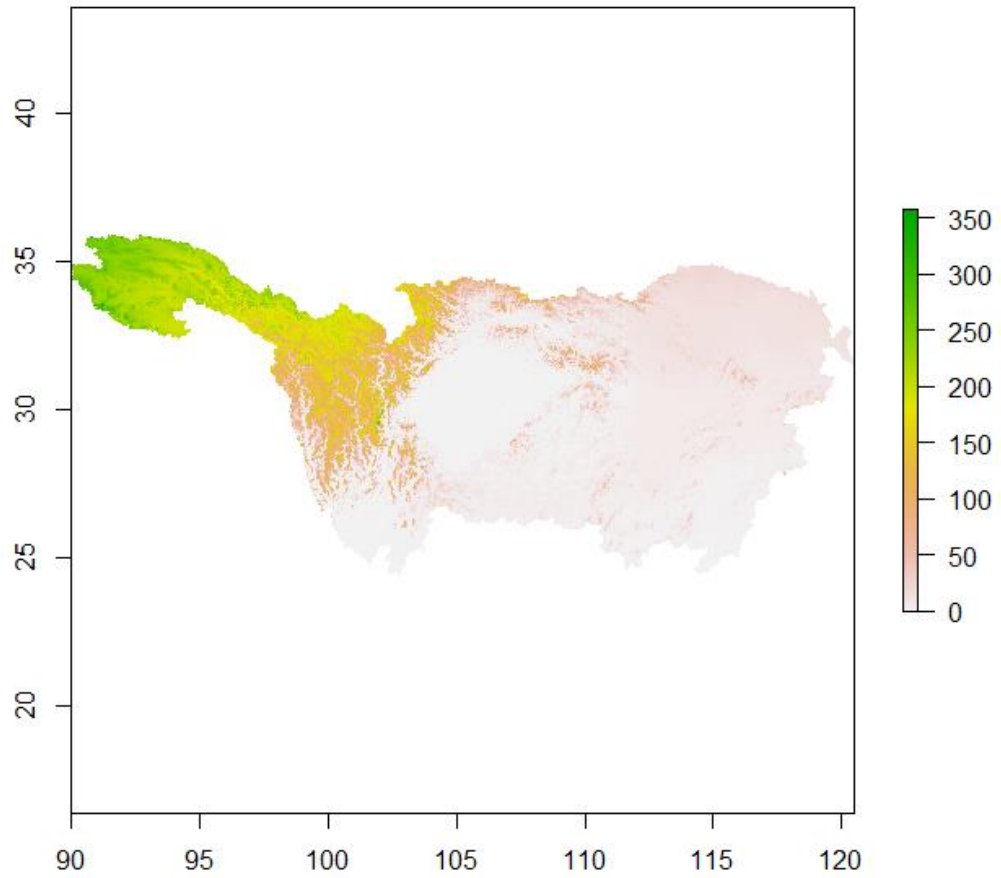


Figure 6.7: Number of frost days across the YDB [units = days]

RETScreen also applies factors to costs to account for varying selling prices of equipment from different manufactures (known as 'K' factor), and local vs Canadian equipment, fuel and labour costs (known as 'B' factor). As this search aims to be independent of manufacturers pricing, the 'K' factor was set to 1. The civil works cost factor 'B' is calculated by:

$$\mathbf{B} = \left(\frac{\mathbf{E_c}}{\mathbf{3}} + \frac{\mathbf{F_c}}{\mathbf{3}}\right) \times \frac{\mathbf{1}}{\left(\frac{\mathbf{E_c}}{\mathbf{L_c}}\right)^{\mathbf{0.5}}} + \frac{\mathbf{1}}{\mathbf{3}}\left(\frac{\mathbf{E_c}}{\mathbf{L_c}}\right)^{\mathbf{0.5}} \mathbf{L_c} \qquad (6.5)$$

where B is the civil works 'B' factor [0.93], E_c is the local vs Canadian equipment cost ratio [1.0], F_c is the local vs Canadian fuel cost ratio [0.94] and L_c is the local vs Canadian labour cost ratio [0.90].

Due to lack of equipment pricing, local equipment is assumed to be the same cost as Canadian equipment and hence the local vs Canadian equipment costs ratio (E_c) is set to 1.0. Despite

the growing economy of China over recent years there is still a difference in labour costs between Canada and China as reflected in figure 6.8 (Trading economics, 2016) with index point units:

Figure 6.8: Graph of Canadian vs China Labour costs (Trading economics, 2016)

$$\therefore \mathbf{L_c} = \frac{\mathbf{104}}{\mathbf{115}} = \mathbf{0.9} \qquad (6.6)$$

Fuel costs also differ between the two countries as fuel costs for 2014 China diesel were 1.09 US $ per litre vs 1.16 US $ per litre for Canada (World Bank, 2017). Hence:

$$\therefore \mathbf{F_c} = \frac{\mathbf{1.09}}{\mathbf{1.16}} = \mathbf{0.94} \qquad (6.7)$$

The ‘F’ and ‘B’ factors are applied to RETScreen cost calculations as determined in table 6.3, dependent on whether the cost applies to ‘local’ or ‘remote’ manufacturing/installations.

Cost	Proportion of cost F factor applied	Proportion of cost B factor applied
C_1, Generator and control	0%	0%
C_2, Turbine	0%	0%
C_3, Installation of energy equipment	100%	100%
C_4, Access road construction	100%	100%
C_5, Transmission line construction	If voltage < 69, 60% If voltage ≥ 69, 40%	If voltage < 69, 60% If voltage ≥ 69, 40%
C_6, Substation and transformer	0%	0%
C_7, Installation of substation and transformer	100%	100%
C_8, Civil works	85%	85%
C_9, Penstock	0%	0%
C_{10}, Installation of penstock	100%	100%
C_{11}, Engineering	40%	0%
C_{12}, Development	0%	0%
C_{13}, Miscellaneous	0%	0%
C_{14}, Feasibility Study	0%	0%

Table 6.3: Application of F and B factors to initial costs

6.5.3 Search the catchment for rivers

Each cell of a raster has a cell number, starting with 1 in the top left-hand corner and the maximum number of cells in the raster at the bottom right. The search starts at a particular start cell number and finishes at an end cell number (set by the user). In this way, the search can be broken into blocks to enable searches in parallel (essential for searching a catchment as large as the YDB). The function extracts the value of the Q_{mean} raster cell by cell, and if a cell has a mean annual flow value of ≥ 10 ls^{-1}, the longitude/latitude coordinates of that cell are extracted. This sets an intake point to investigate. Q_{mean} cells with a value of NA (No data) i.e. non-river, out of catchment or river cells with a mean annual flow of < 10 ls^{-1} are ignored, as is the cell at the catchment outlet. The elevation of the cell is then checked (see section 6.5.4).

6.5.4 Check height function

Using the co-ordinates of the intake point, the elevation of the river is extracted from the DEM. If the height of the river is less than 4000m, the *Check roads* function is initiated, otherwise the intake point is rejected and the search (see section 6.5.3) continues. This cut the number of river points to be examined to 1,378,918 – a reduction of 2.3%. Conditions above 4000m are considered too difficult for hydropower construction and currently the world's highest altitude hydropower site is the newly opened (2015) 510-MW Zangmu Hydropower Station on the Yarlung Zangbo / Brahmaputra River in Tibet, at approximately 3,300m.

6.5.5 Check roads function

The *Check roads* function crops an area of extent 200x200-arcseconds around the intake point and checks if roads are present in that area. If the check returns positive, the investigation continues, otherwise the intake point is abandoned as such remote intake points are unlikely to be viable hydropower sites.

6.5.6 Find turbine data function

In this function, all the input datasets relevant to the search are cropped around the intake point using the 200x200 arc-seconds extent determined in the *Check Roads* function (i.e. DEM, FDR, slope, lakes and reservoirs, available flow, actual flow, number of frost days), which enables the calculations to compute faster. The values of each of these datasets at the intake point are extracted. The lakes and reservoirs dataset is converted from a shapefile to a raster using the *rasterize* function within the *Raster* package, where areas of lake or reservoir have a value of 1. The first turbine point is set as the point on the river one cell distance downstream from the intake point, which is determined by the FDR at the intake point.

The direct distance between the intake point and the local road network is calculated using the *dist2Line* function within the *Geosphere* package, which calculates the shortest distance between points and polylines based on the Harvesine function (a great-circle distance assuming a spherical earth ignoring ellipsoidal effects) where:

$$\mathbf{d} = \mathbf{2r}\ \mathbf{arcsin}\left(\sqrt{\mathbf{sin^2}\left(\frac{\boldsymbol{\varphi_2} - \boldsymbol{\varphi_1}}{\mathbf{2}}\right) + \mathbf{cos}(\boldsymbol{\varphi_1})\,\mathbf{cos}(\boldsymbol{\varphi_2})\,\mathbf{sin^2}\left(\frac{\boldsymbol{\lambda_2} - \boldsymbol{\lambda_1}}{\mathbf{2}}\right)}\right) \qquad (6.8)$$

where d is distance (km), r is the radius of the Earth [6378km] and φ / λ = latitude / longitude of points in radians.

If the distance is less than 1000m, the *dist2Line* function result is accepted. However, if the distance is greater than 1000m, a least-cost path is calculated using the *gdistance* package (see section 6.5.7). Calculating a least-cost path with 2 cells or less would cause the algorithm to fail, hence the 1000m sense check.

Although rivers of equal flow vary in their cross-section, an estimation of the width and depth can be made using power-law relationships with discharge (Andreadis et al., 2013). It is assumed all rivers are single channel rivers. Bank full discharge is estimated to occur between Q_0 and Q_5 (Moody and Troutman, 2002) and hence the river width and depth [m] equations are assumed to be:

$$\mathbf{Width} \approx 7.2\mathbf{Q_5}^{0.5}; \mathbf{Depth} \approx 0.27\mathbf{Q_5}^{0.3} \quad (6.9)$$

6.5.7 Road intake function

A cost grid is calculated based on the DEM crop, where cells with slope greater than 20% and cells that are within lakes or reservoirs given a high cost, and other cells given a low cost. The *gdistance* package *shortestPath* function was used to calculate the least-cost path between two points using a transition matrix conversion of the cost grid based on the function 1/cost. This is because the *gdistance shortestPath* function is based on conductance rather than resistance (see the *gdistance package Vignette* (van Etten, 2017b)). A transition matrix has rows consisting of non-negative real numbers with each row summing to 1, used to describe the transitions of a Markov Chain. Its element in the i^{th} row and j^{th} column describes the probability of moving from state i to state j in one timestep. Error detection code is included, as if this function returns an error, the original *dist2Line* calculation is used.

6.5.8 Check head function

The *Check head* function extracts the elevation data of the newly found turbine point and calculates the gross-head where:

$$\mathbf{Gross\ head}\ (\mathbf{H_g}) = \mathbf{Intake\ elevation} - \mathbf{Turbine\ elevation} \quad (6.10)$$

An initial estimation of the penstock length is calculated by finding the direct 3D distance between the intake and the turbine. If the initial penstock length calculation is below an upper limit (initially set to 30km when testing the code in the Laoguan catchment) and the gross-head (H_g) is greater than 5m and less than 1000m (the upper limit of head for the turbines

shown in figure 6.6) then the *Penstock Length* function is called. If the *Penstock Length* function returns a penstock length between 0 and the upper limit, the *Turbine Intake details* function is subsequently called. Once this is complete, the *Determine turbine point* function is called to search for a new turbine point on the river providing the gross-head and penstock length limits have not been reached.

6.5.9 Determine turbine point function

A new turbine point is found by moving one cell in the FDR direction from the previous turbine point and the *Check head* function is called again.

6.5.10 Penstock Length function

Using a methodology similar to the *Road intake* function, the least cost path between the intake and turbine is found using the *gdistance shortestPath* function. The cost grid is based on the DEM crop, where DEM elevations above the intake height are considered expensive. This is to ensure the penstock is only laid on ground either at the same elevation or lower than the intake height.

The plots in figure 6.9 show both the shortest direct path and the least cost path for an example scheme, and table 6.4 shows the differences in lengths of the penstock using the different methods. Note how the shortest distance passes over ground that is at higher elevation to the intake.

Length by shortest direct route (findDist function)	Length by shortest direct route (gdistance package)	Length by least cost route (gdistance package)
3211m	3231m	3791m

Table 6.4: Penstock lengths of example scheme using different determination methods

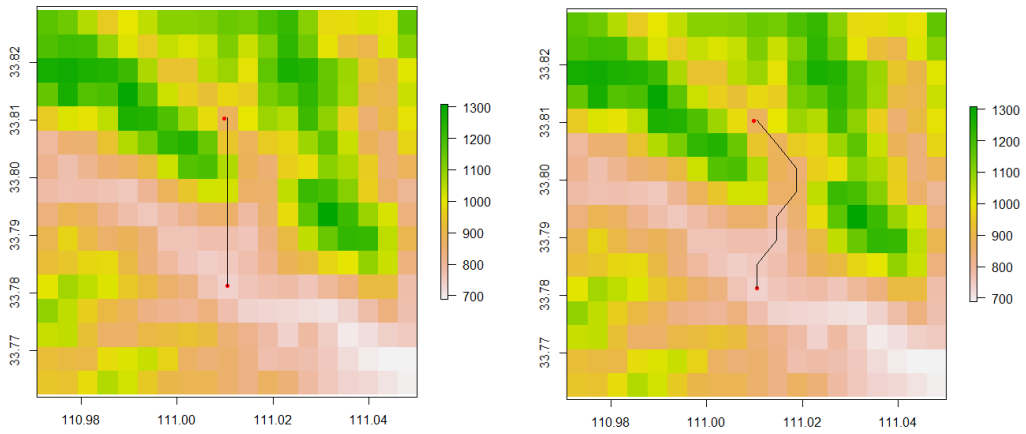


Figure 6.9: Shortest direct path (left) and least cost path (right) of an example penstock (elevations of grid cells in [m])

The *shortestPath* function determines the 2D length, hence the final penstock length is calculated by:

$$\textbf{Penstock Length} = \sqrt{(\textbf{Shortest Path Length}^2 + \textbf{Grosshead}^2)} \qquad (6.11)$$

6.5.11 Turbine Intake details function

Using the same method as the *Road intake* function, the least-cost distance between the turbine and the road network is calculated using the *Road turbine* function, and the least cost road distance between the turbine and the intake is also calculated using the *Intake turbine road length* function. The access road length is then calculated by:

$$\textbf{Acces Road Length} = \min\begin{Bmatrix}\textbf{Turbine to road network} + \textbf{Intake to road network} \\ \textbf{Turbine to road network} + \textbf{Intake to turbine} \\ \textbf{Intake to road network} + \textbf{Intake to turbine}\end{Bmatrix} \qquad (6.12)$$

An initial estimation of the transmission line between the turbine and the 220kV and greater transmission network is made using the *dist2Line* function. The *Iterate design flow* function is then called.

6.5.12 Iterate design flow function

Design flow is the maximum flow of a hydropower site, with any excess flow diverted around the scheme (i.e. excess water does not flow through the penstock). If a very high design flow (e.g. Q_0) is chosen, then the scheme can use all the flow throughout the year, even in flood conditions. However, the turbine efficiency will be low when not at near flood conditions – i.e. the plant capacity factor (the ratio of actual power output to potential output) will be very low. A low design flow (e.g. Q_{90}) on the other hand will only be able to utilise a small amount of the annual river flow but will operate at high capacity as the turbine will be at high efficiency throughout the year. When the river flow is greater than the design flow, the flow remaining once the design flow is abstracted is left in the river i.e. diverted around the penstock. Hence the turbine efficiency does not lower due to having too much flow. Typically turbines drop in efficiency as flow increases above the design flow.

Potential design flows are extracted from the available flow raster stack and, providing the flow is greater than $10ls^{-1}$ ($0.01m^3s^{-1}$), the *Classify project* and *Turbine select* functions are initiated. If these functions return a site classification of at least pico-hydro (i.e. >0.5kW), and a turbine is available for the scheme, the *Initial penstock diameter* function is initiated and then the *Calculate energy* function.

The design flow is iterated, starting at Q_0 in 5% increments and finishing at Q_{90}. Q_{95} and Q_{100} are not available as design flow choices as the 'hands off' flow is Q_{95}.

6.5.13 Classify project function

The *Classify project* function sets an initial classification based on the design flow to estimate the power capacity of the scheme as per table 6.5.

Flow range (m^3s^{-1})	Initial Classification
>50	Large
12.8 to 50	Small
> 0.4 to < 12.8	Mini
< 0.4	Micro

Table 6.5: Initial classification of hydropower schemes (RETScreen, 2004)

It is important to note that different sized schemes can differ significantly in approach and scale, which in turn affects economics.

6.5.14 Turbine select function

Initially setting the number of turbines to 1, the design flow per turbine is calculated by dividing the design flow by the number of turbines and a spatial point on the turbine selection chart is determined by combining this with the gross-head (using the *over* function). Any turbine polygons that the spatial point sits inside are captured, and a 1 or a 0 is assigned if any of the turbine types are available. If at least one turbine type is available, the *Determine Capacity* function is initiated, otherwise the *Add turbine* function is initiated.

6.5.15 Add turbine function

The number of turbines is increased by 1, and providing the number of turbines required is less than 100, the *Turbine select* function is re-run. If the number of turbines required is greater than 100, the algorithm completes and a new design flow choice is selected.

6.5.16 Determine capacity function

The RETScreen equations for determining the power capacity (MW) are initially calculated per turbine and depend on the design flow (in m^3s^{-1}) and gross-head (H_g [m]):

Initial Classification	MW per Unit
Large or Small	$\frac{8.22 * Design\ flow\ per\ Unit * H_g}{1000}$
Mini	$\frac{7.79 * Design\ flow\ per\ Unit * H_g}{1000}$
Micro	$\frac{7.53 * Design\ flow\ per\ Unit * H_g}{1000}$

Table 6.6: Power capacity per unit based on initial classification of hydropower scheme (RETScreen, 2004)

Total capacity is calculated by multiplying the capacity per unit by the number of turbines.

6.5.17 Initial penstock diameter function

Hydraulic losses occur in pipes (i.e. penstocks) as energy is dissipated due to friction, particularly when the flow is turbulent. RETScreen's equation for head loss due to hydraulic losses is calculated by:

$$\mathbf{h_{headloss} = H_g l_{hydr,max} \frac{Q^2}{{Q_d}^2}} \quad (6.13)$$

where $h_{headloss}$ is the hydraulic head-loss [m], H_g is the gross-head [m], $l_{hydr,max}$ is the maximum hydraulic head-loss percentage, Q is the actual flow [m^3s^{-1}] and Q_d is the design flow [m^3s^{-1}].

However, this is a simplification as it does not include the penstock length, and therefore another method was preferred based on the Manning formula (Manning, 1891). It is somewhat surprising that RETScreen does not account for penstock length (Wallace, 2018) as:

$$\mathbf{headloss} \propto \mathbf{f}.\mathbf{L}_p.\frac{v^2}{2g} \quad (6.14)$$

where f is a friction factor, L_p is the penstock length [m], v is the velocity of flow [ms^{-2}] and g is the acceleration due to gravity. As:

$$\mathbf{v = \frac{Q}{A}\ and\ A = \frac{\pi d^2}{4}\ \therefore headloss \propto \frac{L_p}{d^4}} \quad (6.15)$$

Maximum head-loss due to hydraulic losses is assumed to occur at maximum flow and the maximum hydraulic loss can be calculated by the Manning formula for fully closed circular pipe:

$$\mathbf{hf = \frac{10.3 n^2 {Q_d}^2 L_p}{{d_p}^{5.33}}} \quad (6.16)$$

where hf is the head-loss due to friction [m], n is the Manning number for GRP pipe [0.009], and d_p is the penstock diameter [m].

The diameter of the penstock is an important variable in determining head losses due to friction and yet cost also increases with penstock diameter making it an important variable for project economics (Duncan, 2014). Assuming a maximum head-loss of 4%, d_p can be calculated by rearranging the above equations:

$$\mathbf{d_{p\ Manning\ formula} = 0.484\left(\frac{{Q_d}^2 L_p}{H_g}\right)^{0.1876}} \quad (6.17)$$

Note that RETScreen calculates the penstock diameter (d_p) as:

$$\mathbf{d_{p\ RETScreen}} = \frac{\left(\frac{\mathbf{Q_d}}{\mathbf{n_p}}\right)^{\mathbf{0.43}}}{\mathbf{H_g}^{\mathbf{0.14}}} \quad (6.18)$$

where n_p is the number of penstocks. The RETScreen method does not consider the penstock length and therefore may underestimate the starting penstock diameter. The Manning formula method, assuming a maximum headloss of 4%, is therefore a better method for an initial estimation of d_p. However, this only calculates the head loss at the design flow and not across the flow range. Following Duncan's methodology (Duncan, 2014), hf values at lower flows than design flow can be calculated using the Darcy friction factor, f, a dimensionless quantity used in the Darcy-Weisbach equation (Penche, 2004). The average water velocity V within the penstock at the design flow Q_d for a given penstock diameter d_p is:

$$\mathbf{V} = \frac{\mathbf{Q_d}}{\mathbf{A}} = \frac{\mathbf{4Q_d}}{\boldsymbol{\pi}\mathbf{d_p}^{\mathbf{2}}} \quad (6.19)$$

where A is the cross-sectional area of the penstock [m^2]. The Reynolds number R_e is a dimensionless quantity representing the ratio of internal forces to viscous forces, and calculated by:

$$\mathbf{R_e} = \frac{\mathbf{d_p V}}{\boldsymbol{\nu}} \quad (6.20)$$

where v is the kinematic viscosity of water [$1.31*10^{-6} m^2 s^{-1}$]. The Darcy friction factor *f* is calculated by numerically solving the Colebrook White equation:

$$\frac{\mathbf{1}}{\sqrt{\mathbf{f}}} = -\mathbf{2log_{10}}\left(\frac{\boldsymbol{\epsilon}}{\mathbf{3.7d_p}} + \frac{\mathbf{2.51}}{\mathbf{Re}\sqrt{\mathbf{f}}}\right) \quad (6.21)$$

where ϵ is the roughness height [0.029mm for GRP]. In practice, this is solved using the *rootSolve* package (Soetart, 2016) using the function *uniroot*, where the equation above is rearranged to equal 0. The calculated friction factor is worst-case at maximum flow and the head-loss due to friction across the flow range is calculated based upon this worst-case friction factor using the Darcy-Weisbach equation:

$$\mathbf{hf_n} = \mathbf{f}\left(\frac{\mathbf{L_p}}{\mathbf{d_p}}\frac{\mathbf{V_n}^{\mathbf{2}}}{\mathbf{2g}}\right) \quad (6.22)$$

where hf_n is the head-loss at flows less than the design flow [m], g is the acceleration due to gravity [9.81ms^{-2}] and V_n is the water velocity at flows below the design flow [ms^{-1}].

6.5.18 Calculate energy function

The *Calculate energy* function calls subsequent functions to determine the turbine efficiency of any turbines that are available for that flow/gross-head combination.

6.5.19 Turbine efficiency functions

Francis, Kaplan and Propeller type turbines are placed within the flow and turn as fluid flows around them, known as reaction turbines. Pelton, Turgo and Cross-flow type turbines are impulse type turbines, where a fast-moving fluid (or jet) is fired through a narrow nozzle, striking the turbine blades making them spin. All the turbine efficiencies are calculated using the RETScreen methodology.

Common to all reaction type turbines, the turbine runner throat diameter (d_t) is calculated as a function of the design flow (Q_d in m^{3}s^{-1}):

$$\mathbf{d_t = kQ_d^{0.473}} \qquad (6.23)$$

where k is a constant, 0.46 for $d_t < 1.8$m or 0.41 where $d \geq 1.8$m. The software calculates d_t for both constants (0.46 and 0.41) and if both are greater than or equal to 1.8m, then d_t is calculated with k set to 0.41, else it is calculated with k set to 0.46. The specific speed based on flow (n_q) of the turbine is a function of the rated head (h) [m] which is the gross head minus hydraulic losses as calculated in 6.5.17.

$$\mathbf{n_q = kh^{-0.5}} \qquad (6.24)$$

where k is a constant set to 800 for Propeller and Kaplan type turbines or 600 for Francis turbines. To calculate Francis turbine efficiencies, the specific speed based on flow and runner size are adjusted to peak efficiency ($\hat{}e_{nq}$) and ($\hat{}e_d$):

$$\mathbf{\hat{}e_{nq} = \{(n_q - 56)/256\}^2} \qquad (6.25)$$

$$\mathbf{\hat{}e_d = (0.081 + \hat{}e_{nq})(1 - 0.789 d_t^{-0.2})} \qquad (6.26)$$

The turbine peak efficiency e_p includes a manufacturer/design coefficient (R_m) which is set to the default value of 4.5:

$$e_p = (0.919 - \hat{}e_{nq} + \hat{}e_d)\text{-}0.0305\text{+}0.005R_m \quad (6.27)$$

The peak efficiency flow (Q_p) and efficiencies below peak efficiency flow ($e_{q\ below}$) are found by:

$$Q_p = 0.65 * Q_d * n_q^{0.05} \quad (6.28)$$

$$e_{q\ below} = \left\{1 - \left[1.25\left(\frac{Q_p - Q}{Q_p}\right)^{(3.94-0.0195n_q)}\right]\right\} e_p \quad (6.29)$$

When at full load, there is a drop-in efficiency ($\hat{}e_p$) which is used to calculate the efficiency at full load (e_r) and efficiencies at flows above peak efficiency flow ($e_{q\ above}$):

$$\hat{}e_p = 0.0072n_q^{0.4} \quad (6.30)$$

$$e_r = (1 - \hat{}e_p)e_p \quad (6.31)$$

$$e_{q\ above} = e_p - \left[\left(\frac{Q - Q_p}{Q_d - Q_p}\right)^2 (e_p - e_r)\right] \quad (6.32)$$

Note that peak efficiency flow is less than design flow and there is a drop in efficiency when at design flow compared to peak flow. However, although a turbine would continue to drop in efficiency if flows were allowed above the design flow, any excess flow above design flow is diverted around the hydropower scheme, via a spillway for example, as flow control valve(s) will limit flow through the turbine. Therefore, once the flow is at or above design flow, the efficiency is constant and this applies to all turbine types considered.

Like the Francis turbine, the specific speed based on flow and runner size of Kaplan and Propeller turbines are adjusted to the peak efficiency ($\hat{}e_{nq}$) and ($\hat{}e_d$) and the peak efficiency (e_p) calculated as follows:

$$\hat{}e_{nq} = \{(n_q - 170)/700\}^2 \quad (6.33)$$

$$\hat{}e_d = (0.095 + \hat{}e_{nq})(1 - 0.789d^{-0.2}) \quad (6.34)$$

$$\mathbf{e_p = (0.905 - \hat{}e_{nq} + \hat{}e_d)\text{-}0.0305\text{+}0.005R_m} \quad (6.35)$$

For the Kaplan turbine, the peak efficiency flow (Q_p) and efficiencies below and above peak efficiency flow (e_q) are found by:

$$\mathbf{Q_p = 0.75 * Q_d} \quad (6.36)$$

$$\mathbf{e_q = \left\{1 - \left[3.5\left(\frac{Q_p - Q}{Q_p}\right)^{1.13}\right]\right\} e_p} \quad (6.37)$$

The propeller turbine peak efficiency flow (Q_p) and efficiencies below and above peak efficiency flow (e_q) are found by:

$$\mathbf{Q_p = Q_d} \quad (6.38)$$

$$\mathbf{e_q = \left\{1 - \left[1.25\left(\frac{Q_p - Q}{Q_p}\right)^{1.13}\right]\right\} e_p} \quad (6.39)$$

For Pelton turbines, the rotational speed (n) and outside diameter of the runner (d_r) is calculated by:

$$\mathbf{n = 31\left(\frac{Q_d}{j}\right)^{0.5}} \quad (6.40)$$

$$\mathbf{d_r = \frac{49.4 h^{0.5} j^{0.02}}{n}} \quad (6.41)$$

where j is the number of jets and h is the rated head [m]. It is also somewhat surprising that the rotational speed (n) does not include a term to represent h_{net} (Wallace 2018) as the rotational runner speed is normally proportional to h_{net} as:

$$\boldsymbol{n \propto \frac{\sqrt{h_{net}}}{d_p}} \quad (6.42)$$

Although the number of jets can vary, here j is set to 3 to reduce the number of necessary iterations of design. The turbine peak efficiency (e_p), peak efficiency flow (Q_p) and efficiency at flows above and below peak efficiency flow (e_q) are calculated as:

$$\mathbf{e_p = 0.864d^{0.04}} \quad (6.43)$$

$$\mathbf{Q_p = (0.662 + 0.001j)Q_d} \quad (6.44)$$

$$\mathbf{e_q = \left\{1 - \left[(1.31 + 0.025j)\left|\frac{Q_p - Q}{Q_p}\right|^{(5.6+0.4j)}\right]\right\}e_p} \quad (6.45)$$

Turgo turbine efficiencies are calculated as:

$$\mathbf{e_{q\ Turgo} = e_{q\ Pelton} - 0.03} \quad (6.46)$$

For the cross-flow turbine, the peak efficiency flow and efficiencies below and above peak are calculated as:

$$\mathbf{Q_p = 0.75 * Q_d} \quad (6.47)$$

$$\mathbf{e_{q,crossflow} = 0.79 - 0.15\left(\frac{Q_d - Q}{Q_p}\right) - 1.37\left(\frac{Q_d - Q}{Q_p}\right)^{14}} \quad (6.48)$$

6.5.20 Turbine power calculation functions

For each percentile flow, the turbine efficiencies are calculated with the flow set to the minimum of the design flow and actual river flow, together with any head loss as set by the equations in 6.5.17. The tail water effect is the loss of head due to rising river levels below the turbine outlet, calculated in this work by Andreadis et al. (2013) river depth equations (see p82):

$$\mathbf{If\ River\ flow > Design\ flow,}$$
$$\mathbf{Tailwater\ effect = 0.27 * \left(Riverflow^{0.3} - Designflow^{0.3}\right)} \quad (6.49)$$
$$\mathbf{else, Tailwater\ effect = 0}$$

The power produced for each percentile flow is calculated by RETScreen as follows:

$$\mathbf{P_n = \rho g Q\left[H_g - \left(h_{hydr} - h_{tail}\right)\right]e_t e_g (1 - l_{trans})(1 - l_{para})} \quad (6.50)$$

where P_n is the power at percentile flow Q_n [kW], Q_n is the percentile flow [m^3s^{-1}], ρ is the density of water = 1000 [kgm^3s^{-1}], g is the acceleration due to gravity [$9.81ms^{-2}$], H_g is the gross-head [m], h_{hydr} are hydraulic losses, h_{tail} is the head-loss due to the tail water effect, e_t is

the turbine efficiency, e_g is the generator efficiency assumed to be 0.98, l_{trans} are transformer losses and l_{para} are parasitic losses (the latter two assumed to be 2%).

The final plant capacity (P[kW]) equation can be calculated by:

$$\mathbf{P\ [kW] = \rho g\ (1 - h_{hydr}) * Q_d H_g e_{tQd} e_g (1 - l_{trans})(1 - l_{para})} \qquad (6.51)$$

where e_{tQd} is the turbine efficiency at design flow and h_{hydr} is assumed to be 4%.

The energy production potential is determined by calculating the area under the power duration curve assuming a straight line between adjacent calculated power output values. The power duration curve represents an annual cycle, and each 5% interval on the curve is equivalent to 5% of 8760 hours (i.e. the number of hours in a year). Using the power values P (in kWh), the energy production potential (E) in kWhr yr^{-1} is calculated using:

$$\mathbf{E} = \sum_{\mathbf{k=1}}^{\mathbf{20}} \left(\frac{\mathbf{P_{5(k-1)} + P_{5k}}}{\mathbf{2}} \right) \frac{\mathbf{5}}{\mathbf{100}} \mathbf{8760(1 - l_{dt})} \qquad (6.52)$$

where P_5 is the power at 5% percentile flow intervals and l_{dt} is the percentage loss due to downtime (e.g. maintenance etc), set to 2%. The plant capacity factor is a measure of how efficient the plant is and defined as the output of the plant compared to its rated capacity:

$$\mathbf{Plant\ capacity\ factor} = \frac{\mathbf{E}}{\mathbf{P * 8760}} \qquad (6.53)$$

6.5.21 Transmission Line Voltage function

As discussed in 3.9, an electrical grid connection dataset was generated from a GIS map of 220kV and greater transmission lines and a proxy dataset consisting of locations of villages, towns and cities. Based on the classification of each hydropower configuration (see *table 6.7*), the distance between the turbine and the nearest village, town or city was found.

Final Classification	Possible connections
Large	City or 220kV and greater network
Small	Town or City
Mini	Town or City
Micro	Village, town or city

Table 6.7: Transmission network connections for each classification of hydropower scheme

Distance between the turbine and nearest village/town/city was established by first creating a data table based on the village/town/city shapefile, listing each location's latitude and longitude and whether it is a village/town/city. The nearest village/town/city to the turbine was found using the *get.knnx* function within the *FNN* package, enabling a fast k-nearest neighbour search. Longitude and latitude of the nearest neighbour was then converted to radians and the distance between the turbine and village/town/city established by:

$$\boldsymbol{Distance = 2 * r} \\ \boldsymbol{* \arcsin\left(sin\left(\frac{\varphi_{VTC} - \varphi_t}{2}\right)^2 + \left(cos(\varphi_t)\cos(\varphi_{VTC})\sin\left(\frac{\lambda_{VTC} - \lambda_t}{2}\right)\right)\right)} \quad (6.54)$$

where r is the Earth's radius [6371km], φ_{VTC} is the latitude of nearest village/town or city [radians], φ_t is the latitude of the turbine , λ_{VTC} is the longitude of nearest village/town or city [radians] and λ_t is the longitude of the turbine [radians].

The transmission network within this model comprises the transmission voltages of 11kV, 33kV, 66kV, 110kV, 132kV, 200kV, 220kV, 330kV, 500kV, 800kV, 1000kV and 1100kV and power from all potential schemes is transmitted at the lowest possible voltage. The voltage requirements of the transmission network can be calculated by rearranging:

$$\mathbf{P = \sqrt{3} V I \cos\phi} \quad (6.55)$$

where P is the power or plant capacity [W], V is the transmission line voltage [V], I is the current [Amperes] and cos ϕ is a power factor which is assumed to be 0.8. However, cos ϕ will vary dependent on the particular system setup.

Current carrying capacity values have been estimated from the Indian company Uttar Pradesh Power Transmission Corporation Limited (Uttar Pradesh Power Transmission Corporation Limited, 2016) for four aluminium conductor steel-reinforced cable conductors used in the Uttar Pradesh power network at 65°C (see table 6.8):

Conductor Name	Voltage	New conductor current carrying capacity (amperes)	Old conductor current carrying capacity (amperes)	Mid-point current carrying capacity (amperes)
Dog	<= 132kV	141.12	150.20	145.66
Panther	132kV to 220kV	179.89	200.60	190.25
Zebra	330kV	201.26	249.51	225.39
Moose	> 400kV	133.60	218.89	176.25

Table 6.8: Voltage range and current carrying capacity of four aluminium conductor steel-reinforced cable conductors (Uttar Pradesh Power Transmission Corporation Limited, 2016)

The lowest acceptable voltage conductor type is selected, and the required voltage set based on that conductor type. Transmission line voltage is then set by rounding up the required voltage to one of the transmission voltages on the Chinese network. Finally, the distance between each turbine to the 220kV and greater network is calculated using the *dist2line* function (discussed earlier). Transmission line distances for plants with a voltage requirement of 330kV are estimated between the turbine and 330kV and greater network, and those with a voltage requirement of >=500kV estimated between the turbine and the 500kV and greater network. The final transmission line distance is the minimum of the distance to the nearest appropriate village/town/city or to the appropriate 220kV and greater network.

A simpler, alternative method may be considered for future development. By assuming an upper limit of 2000A circuit breakers and knowledge of the power factor, the maximum power transmitted can be calculated for each voltage network (Wallace, 2018).

6.5.22 *Costing* function

RETScreen costing formulae described in this section include the civil works factor B and frost factor F as discussed in 6.5.2 and weighted according to table 6.3. All costs are given in Chinese Yuan (CNY).

RETScreen assumes the generator and control is standard for all turbine types. Generators convert the mechanical energy of the water into electrical energy and the cost (C_1) equated by:

$$\mathbf{C_1 = 0.82\ n^{0.96} GC_g \left(\frac{MW}{H_g^{0.28}}\right)^{0.9} \times 10^6 \times C_{con}} \qquad (6.56)$$

where n is the number of turbines, G is a grid connection factor, set to 0.9 if MW < 1.5 and central-grid connected else set to 1.0, C_g is a lower cost motor factor where C_g is set to 0.75 if MW < 10 else C_g is set to 1.0, MW is the plant capacity [MW], H_g is the gross-head [m] and C_{con} is a conversion from RETScreen 2004 Canadian dollars cost to Chinese Yuan [8.975].

RETScreen lumps the governor cost together with the turbine cost and for reaction turbines (Francis, Kaplan and Propeller) each have a different cost formula. For Kaplan turbines, the cost (C_2) is:

$$\mathbf{C_{2,Kaplan} = 0.27\ n^{0.96} J_t\ K_t\ d^{1.47}\left(1.17\ H_g{}^{0.12} + 2\right) \times 10^6 \times C_{con}} \qquad (6.57)$$

where J_t is a higher cost vertical axis turbine factor and set to 1.0 if $H_g \leq 25$m, else set to 1.1, K_t is a lower cost small horizontal axis turbine factor and set to 0.9 if d < 1.8, else set to 1.0, and d is the runner diameter [m].

Francis and Propeller types turbines are costed as:

$$\mathbf{C_{2,Francis} = 0.17\ n^{0.96} J_t\ K_t\ d^{1.47}\left\{\left(13 + 0.01 H_g\right)^{0.3} + 3\right\} \times 10^6\ \times C_{con}} \qquad (6.58)$$

$$\mathbf{C_{2,Propeller} = 0.125\ n^{0.96} J_t\ K_t\ d^{1.47}\left(1.17\ H_g{}^{0.12} + 4\right) \times 10^6 \times C_{con}} \qquad (6.59)$$

Pelton and Turgo turbine types are assumed to be equal cost and the Crossflow turbine assumed to be half that of the Pelton/Turgo cost, with the formula selected dependent on the ratio of capacity per turbine to gross-head:

$$\mathbf{if\ \frac{MW_u}{H_g{}^{0.5}} > 0.4, C_{2(Pelton/Turgo)} = 3.47\ n^{0.96}\left(\frac{MW_u}{H_g{}^{0.5}}\right)^{0.44} \times 10^6 \times C_{con}} \qquad (6.60)$$

$$\mathbf{if\ \frac{MW_u}{H_g{}^{0.5}} \leq 0.4, C_{2(Propeller)} = 5.34\ n^{0.96}\left(\frac{MW_u}{H_g{}^{0.5}}\right)^{0.91} \times 10^6 \times C_{con}} \qquad (6.61)$$

$$\mathbf{C_{2(Crossflow)} = 0.5\ \times C_{2(Pelton,Turgo)} \times 10^6 \times C_{con}} \qquad (6.62)$$

where MW_u is the capacity per turbine [MW].

Installation of energy equipment costs (C_3) are related to the generator/control and turbine costs but do not include the C_{con} conversion factor as this has already been used in generating costs C_1 and C_2:

$$\mathbf{C_3\ = 0.15\ (C_1 + C_2) \times F \times B} \qquad (6.63)$$

It is assumed that all access roads are unpaved roads (labelled as Tote roads in RETScreen), and the RETScreen formula of costing access roads (C_4) is:

$$\mathbf{C_4 = 0.025TA^2{l_a}^{0.9} \times 10^6 \times C_{con} \times F \times B} \quad (6.64)$$

where T is a tote road factor (set to 0.25 for all projects), A is a road difficulty factor (set to 2 for all roads) and l_a is the length of access road [km].

The RETScreen equation for the transmission line cost (C_5) is as follows:

$$\mathbf{C_5 = 0.0011DP{l_T}^{0.95}V \times 10^6 \times C_{con} \times (FBT_{local} + (1 - T_{local}))} \quad (6.65)$$

where D is the transmission line installation difficulty factor (set to 1.5), P is a factor to reflect cost of wood pole vs steel tower construction (P set to 0.85 if V < 69KV, otherwise P is 1.0), l_T is the length of transmission line [km], V is the voltage [kV] and T_{local} is a local installation factor (assumed to be 0.6 if V < 69kV, otherwise 0.4).

RETScreen formula for substation and transformer cost (C_6) is:

$$\mathbf{C_6 = 0.0025n^{0.95} + 0.002(n+1) \times \left(\frac{MW}{0.95}\right)^{0.9} \times V^{0.3} \times 10^6 \times C_{con}} \quad (6.66)$$

where n is the number of turbines, MW is the plant capacity [MW] and V is the transmission voltage [kV].

Installation costs are related to C_6:

$$\mathbf{C_7 = 0.15C_6 \times F \times B} \quad (6.67)$$

Civil works include construction of the weir, intake works, power house (turbine house) and tailrace which takes water from the turbine back to the river. The RETScreen equation for civil works costs (C_8) is dependent on the size of the scheme.

If classification is 'small' or 'large':

$$\mathbf{C_8 = 3.54n^{-0.04}CR \times \left(\frac{MW}{{H_g}^{0.3}}\right)^{0.82} \times (1 + 0.01l_b) \times \left(1 + 0.005\frac{l_d}{H_g}\right) \times 10^6 \times C_{con} \times (0.85FB + 0.15)} \quad (6.68)$$

If classification is 'mini':

$$\mathbf{C_8 = 3.54n^{-0.04}CR \times \left(\frac{MW}{H_g{}^{0.3}}\right)^{0.82} \times (1 + 0.01l_b) \times \left(1 + 0.005\frac{l_d}{H_g}\right) \times 10^6 \times C_{con} \times (0.85FB + 0.15)} \quad (6.69)$$

If classification is 'micro':

$$\mathbf{C_8 = 1.97n^{-0.04}C \times \left(\frac{MW}{H_g{}^{0.3}}\right)^{0.82} \times \left(1 + 0.005\frac{l_d}{H_g}\right) \times 10^6 \times C_{con} \times (0.85FB + 0.15)} \quad (6.70)$$

where n is the number of turbines, C is a civil cost factor depending if an existing dam is in place which are not accounted for and therefore C is set to 1, R is a rock factor (1.05 if rock is in place at construction site, otherwise 1.0 and set to be 1.05, l_b is the distance to borrow pits set to 0.5km for all sites [km], l_d is the dam crest length [m], set to $7.2Q_5{}^{0.5}$. The dam crest length is assumed to be the calculated width of the river at near flood conditions (i.e. actual flow Q_5).

RETScreen penstock costs (C_9) assumes the penstock is constructed from steel, potentially making the penstock a particularly expensive component of the system. Modern day penstocks are constructed from GRP pipe due to the cost benefits over steel, as the penstocks are both lighter and raw material costs lower. RETScreen uses the weight of the penstock material in the costing equation and the same methodology applied for calculation of GRP pipe costs. The calculation for the weight (W_{steel}) of the steel pipe is:

$$\mathbf{W_{steel} = 24.7 d_p l_p t_{av}} \quad (6.71)$$

where d_p is the diameter of the penstock [m], l_p is the length of the penstock [m] and t_{av} is the average thickness of penstock walls [mm], and calculated by:

$$\mathbf{t_{av} = 0.5(t_t + t_b), \text{ where } t_t = d_p{}^{1.3} + 6\ [mm] \text{ and } t_b = 0.0375 d_p H_g\ [mm]} \quad (6.72)$$

where t_t is the thickness at the top of the penstock [mm] and t_b is the thickness at the base [mm]. Density of the GRP penstock material is approximately one quarter of steel (EBS, 2016) and raw material costs are approximately 55% of steel (Fibrolux, 2016). Therefore, the weight of GRP (W_{grp}) pipe penstock is calculated by:

$$\mathbf{W_{grp} = 6.5 d_p l_p t_{av}} \quad (6.73)$$

The final costing equation for steel pipe from RETScreen ($C_{9\ steel}$) is:

$$\mathbf{C_{9\ steel} = 20 n_p^{0.95} W_{steel}^{0.88} \times C_{con}} \quad (6.74)$$

where n_p is the number of penstocks. Hence costing using GRP pipe ($C_{9\ grp}$) is equal to:

$$\mathbf{C_{9\ grp} = 0.55 \times 20 n_p^{0.95} W_{grp}^{0.88} \times C_{con}} \quad (6.75)$$

Installation cost of the penstock (C_{10}) is also related to the weight of the material:

$$\mathbf{C_{10} = 5 W_{grp}^{0.88} \times C_{con} \times FB} \quad (6.76)$$

Canals and tunnels constructed at the intake point of a river can be used to divert water away from the natural river path, possibly to other rivers or across hill slopes to move the intake point nearer to the power house whilst still maintaining head. This can reduce civil works and/or penstock costs. There are many permutations involved where canals/tunnels are considered and hence difficult to automate, so in this project they are omitted. However, it is worth bearing in mind that canal/ tunnel costs are low relative to penstock costs and they may offer a significant economic benefit if considered.

RETScreen includes other costs to account for engineering, miscellaneous, development and feasibility studies. Engineering costs (C_{11}) include the detail design and project management:

$$\mathbf{C_{11} = 0.37 n^{0.1} E \left(\frac{MW}{H_g^{0.3}} \right)^{0.9} \times 10^6 \times C_{con} \times (0.4F + 0.6)} \quad (6.77)$$

where n is the number of turbines, E is an engineering cost factor to account for existing dams in projects (set to 1.0 as no existing dams accounted for), MW is the plant capacity [MW] and H_g is the gross-head [m].

It is not clear exactly what development costs (C_{12}) include but assumed to be the necessary steps to ensure the project happens (permits, liaison with interested parties etc) and is a proportion of the costs already discussed:

$$\mathbf{C_{12} = 0.04 \sum (C_1 \text{ to } C_{11})} \quad (6.78)$$

Miscellaneous costs (C_{13}) are a function of the interest rate % (i), the design flow and a proportion of all the other costs and depend on the size of the scheme:

For mini, small and large projects:

$$\mathbf{C_{13} = 0.25iQ_d^{0.35} \times 1.1 \sum (C_1 \text{ to } C_{12}) + 0.1 \sum (C_1 \text{ to } C_{12})} \quad (6.79)$$

For micro-hydro projects:

$$\mathbf{C_{13} = 0.17i \times 1.1 \sum (C_1 \text{ to } C_{12}) + 0.1 \sum (C_1 \text{ to } C_{12})} \quad (6.80)$$

where i is the interest rate (assumed to be 6%) and Q_d is the design flow [m^3s^{-1}].

Feasibility study costs (C_{14}) are again a proportion of the total other costs and depend on the size of the scheme. For mini, small and large projects:

$$\boldsymbol{C_{14} = 0.032 \sum (C_1 \text{ to } C_{13})} \quad (6.81)$$

For micro-hydro projects:

$$\boldsymbol{C_{14} = 0.031 \sum (C_1 \text{ to } C_{13})} \quad (6.82)$$

Total initial (up front) costs (C_0) are a summation of all the cost values found thus far:

$$\mathbf{C_0 = \sum (C_1 \text{ to } C_{14})} \quad (6.83)$$

Operation and maintenance costs are set to 3% of the total initial costs over the project life cycle (25 years):

$$\textbf{Operation \& maintenance costs (over 25y)} = \textbf{Total initial costs} \times 3\% \quad (6.84)$$

Hence:

$$\textbf{Total costs (over 25 years)} = 1.03 * \textbf{Total initial costs} \quad (6.85)$$

The costing function also calculates the project economics (see section 6.6) and adds all the data of the current search to the results data table.

6.6 Project Economics

Electricity price in China in 2015 was reported as between 11-12 US cents per kWh (Statista, 2017), equating to between 680 and 742 CNY per MWh. It is assumed that the energy price per MWh is between these two figures (722 CNY per MWh) and all the energy produced is sold. Therefore, the total revenue per annum is:

$$\textbf{Annual revenue} = (\textbf{EI}) - \textbf{O\&M} \quad (6.86)$$

where E is the annual energy produced [MWh], I is the income per MWh [722CNY] and O&M are operation and maintenance costs.

Installed costs per kW [CNY kW^{-1}] are:

$$\textbf{Installed costs} = \frac{\mathbf{C_t}}{\mathbf{P}} \tag{6.87}$$

where C_t are the total investment costs [CNY] and P is the plant capacity [kW].

Net Present Value (NPV) is used in capital budgeting to analyse the profitability of a projected investment of a project, and is the difference between the present value of cash inflows and the present value of cash outflows, equated by:

$$\mathbf{NPV} = \sum_{\mathbf{t=1}}^{\mathbf{t=n}} \frac{\mathbf{C_t}}{(\mathbf{1}+\mathbf{r})^{\mathbf{t}}} - \mathbf{C_0} \tag{6.88}$$

where t is the time-period [years], n is the lifetime of the project, set to 25 [years], r is the discount rate [%] and C_t is the net cash inflow during the period t;

Each year is assumed to generate the same income/costs and the NPV calculated for discount rate scenarios of 5%, 10% and 15%. Discount rate considers the interest rate on loans and the risk and uncertainty of future cash flows. NPV was calculated using the *npv* function within the *FNN* package. A scheme with an NPV of 0 or above at a particular discount rate is viewed as a viable scheme.

The discounted payback period (DPP) was calculated to show clearly the payback timescale:

$$\boldsymbol{DPP} = \boldsymbol{ln}\left(\frac{\mathbf{1}}{\mathbf{1}-\frac{\boldsymbol{C_0 r}}{\boldsymbol{C_t}}}\right) \div \mathbf{ln}(\mathbf{1}+\boldsymbol{r}) \tag{6.89}$$

6.7 Testing the hydropower search algorithm to optimise the search parameters

With such an extensive river network within the YDB, running the hydro search algorithm without constraints could lead to a search that takes excessively long, including searching hydropower scheme configurations that would be unlikely to be funded in the real world. Therefore, the algorithm was tested on a test catchment (Laoguan He) with view to establishing parameter ranges of unviable schemes to reduce the search size. This was initially

run without constraints, testing all flow conditions as potential design flows limited only by a penstock maximum length of 30km and maximum head of the turbine map in figure 6.6 (1000m).

The search produced a data table consisting of 3,834,592 configurations of which only 169,828 had an IRR >= 5%. Input characteristics of this reduced IRR >=5% dataset included:

Input parameter	**Parameter range of viable schemes (schemes with IRR >= 5%)**
Design Flow	> $0.02m^3s^{-1}$ [> $20ls^{-1}$]
Q Selection	Q_5 to Q_{85}
Classification	Micro to large (pico not viable)
Plant Capacity	> 10kW
Grosshead	10m to 1000m
Turbine types	All
Penstock distance	386m to 30km

Table 6.9: Input parameter range of viable hydropower configurations of the Laoguan catchment

Applying these constraints to the full Laoguan results table reduces the size of the search from 3,834,592 to 1,372,696, an initial reduction of ~65%. However, there are still many permutations tested due to the potentially long penstock distance and many iterations of Q. Maintenance and flow issues aside, installing one long penstock within a hydropower scheme may be profitable for a developer of a single scheme due to the increased head (and hence power). This would be a benefit as civil works costs of a scheme would be similar to one with a shorter penstock and less head. However, this is not necessarily the best configuration if the aim is to maximise the NPV or energy output from a whole catchment perspective.

Although 169,828 configurations were deemed viable (i.e. IRR>=5%), not all of these could be installed within the Laoguan catchment as they would conflict with each other (only one scheme can be placed within one stretch of river). Therefore, a hydropower conflict algorithm (see section 6.8) was developed to select the best performing schemes for the catchment that do not conflict with other schemes. This conflict algorithm was also used to show that limiting the penstock length was also beneficial for maximising NPV or energy output from the catchment whilst also limiting the number of searches performed by the hydropower search algorithm.

6.8 Developing a hydropower conflict algorithm

Using the viable results dataset (IRR >=5%), configurations with the same intake and turbine location were grouped together and the scheme with the best NPV at 5% discount rate was selected, achieved by using the *ddply* function of the *plyr* package. *ddply* applies a function (in this case maximum NPV) to each subset group and combines the outputs into a new table referred to as *Best NPV*. For the Laoguan catchment this resulted in a table of 11,033 configurations, with the same penstock length range as the previous dataset, but now with only Q selection design flows between Q_{10} and Q_{50}.

The hydropower conflict algorithm operates as follows:

1) Start by creating an empty data table of best performing, conflict free configurations (referred to as *Accepted results*).
2) The configuration with the maximum NPV is selected from the *Best NPV* table and the *sense check* function is initiated (step 4) – if *Best NPV* is empty go to step 14.
3) Note if the *Accepted results* table is empty (which it is at the beginning of the algorithm), the configuration with the maximum NPV is automatically copied across to *Accepted results*.
4) The *sense check* function sets the current river location to the intake point of the configuration with the maximum NPV found in step 2.
5) The current location is moved one cell down the river in the direction of the flow direction raster (FDR).
6) If there is a match between the current river location and any intake or turbine locations in the *Accepted results* table, then the selected configuration with the maximum NPV conflicts with an existing accepted site (go to step 8), otherwise go to step 7.
7) Step 5 and 6 are repeated until either a conflict is found or the river location is at the turbine location of the configuration with the maximum NPV.
8) If there is a conflict, this configuration is not copied to the *Accepted results* table and removed from *Best NPV* - return to step 2.
9) If there is no conflict, then the configuration is accepted and copied to the *Accepted results* table.

10) The current river location is set to the intake of the accepted configuration and any rows in the *Best NPV* data table with the same intake or turbine location as the current river location are removed.
11) The current location is moved one cell down the river in the direction of the flow direction raster (FDR).
12) If there is a match between the current river location and any intake or turbine locations in table *Best NPV*, then these rows are removed from *Best NPV* and returns to step 11 until the current location is at the turbine of the accepted configuration (see step 13).
13) When at the turbine location of the accepted configuration, any configurations in *Best NPV* with the same intake or turbine as the turbine of the accepted configuration are removed, and the algorithm returns to step 2.
14) When the *Best NPV* table is empty, the algorithm terminates.

Using the Laoguan catchment as an example, the conflict free hydropower schemes (i.e. those in the *Accepted results* table) at 5% discount rate (with the maximum penstock length set to 30km) gives a total installed capacity of 140MW and 356GWhryr^{-1}. This is generated from 21 ROR hydropower stations with an NPV at 5% discount rate of 1.416 billion CNY (see figure 6.10). However, if the maximum penstock length is restricted to 10km and the hydropower conflict algorithm re-run, the installed capacity increases to 144MW generating 362GWhr yr^{-1}. This enables 22 ROR hydropower stations, with an NPV at 5% of 1.609 billion CNY (see figure 6.11). Table 6.10 demonstrates how penstock length restriction affects catchment wide installed capacity, energy generation and NPV at 5% discount rate.

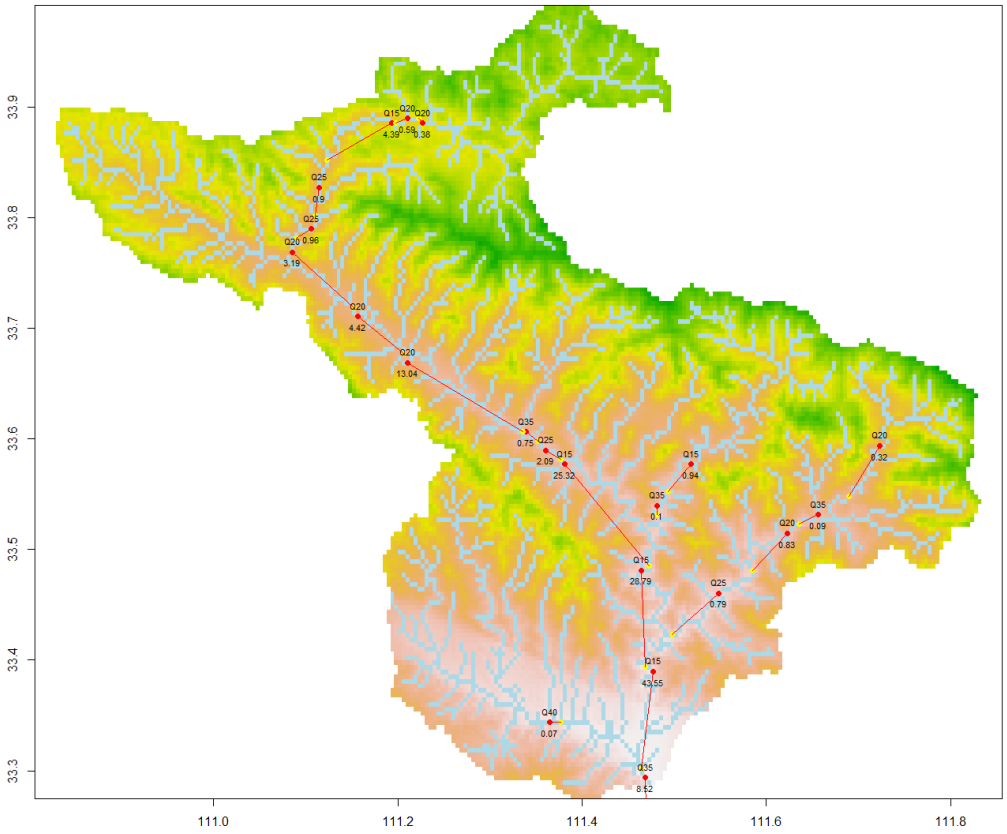


Figure 6.10: Conflict free run-of-river hydropower stations in the Laoguan catchment with maximum penstock length restricted to 30km (installed capacity of 140MW)

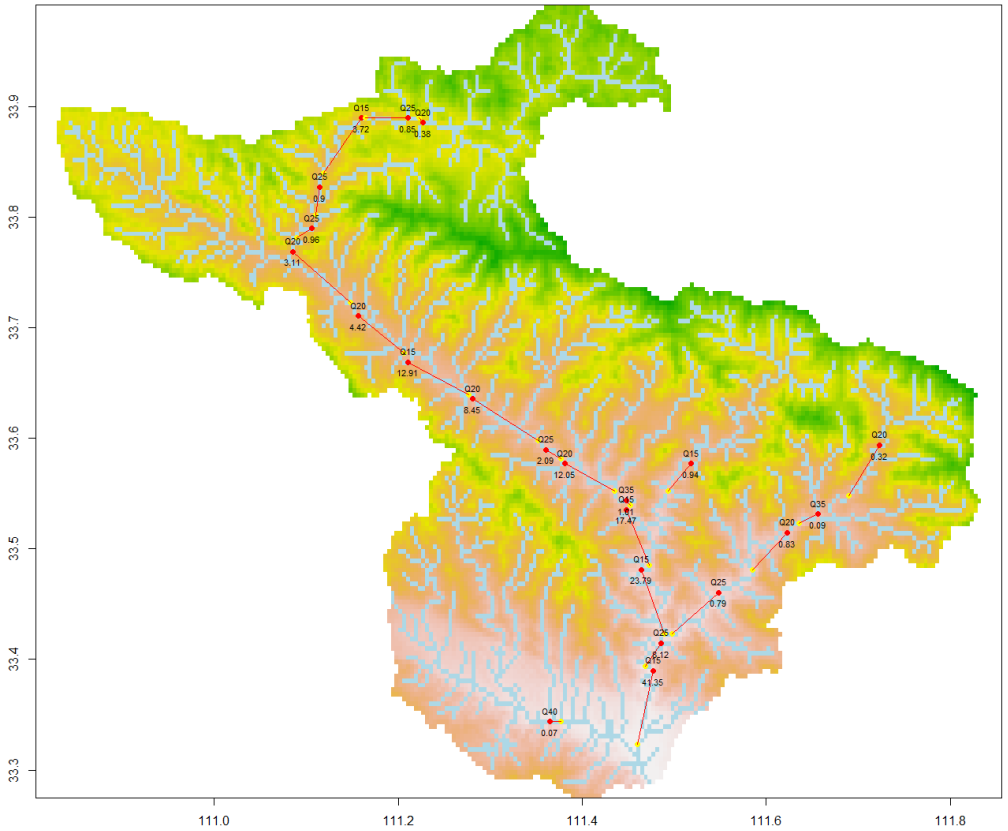


Figure 6.11: Conflict free run-of-river hydropower stations in the Laoguan catchment with maximum penstock length restricted to 10km (installed capacity of 144MW)

Table 6.10 demonstrates that maximum NPV is generated by restricting the penstock length to 10km, but more energy is generated by restricting the penstock length to 13km. It was therefore decided to limit penstock length to 15km when running the hydropower search across the YDB which would at least halve the search time compared with allowing a maximum penstock length of 30km.

Maximum penstock length restriction [km]	Laoguan catchment installed capacity [MW]	Number of run-of-river stations	Energy generation [GWhr yr^{-1}]	NPV at 5% discount rate
30	140	21	356	1.416 billion
25	141	21	359	1.410 billion
20	137	21	343	1.366 billion
15	139	22	353	1.486 billion
14	139	22	353	1.486 billion
13	147	23	372	1.563 billion
12	146	22	368	1.558 billion
11	144	21	361	1.603 billion
10	145	22	362	1.610 billion
9	140	22	351	1.576 billion
8	147	25	366	1.603 billion
7	139	29	358	1.587 billion
6	136	28	350	1.566 billion
5	136	30	358	1.560 billion

Table 6.10: Impact of maximum penstock length on installed capacity, number of stations, energy generation and NPV (5%) in the Laoguan catchment

6.9 Final hydropower search parameter range

With view to limiting the YDB search to those that are most likely to be viable, the final parameter search ranges are shown in table 6.11.

Input parameter	Parameter range of viable schemes (schemes with IRR >= 5%)
Design Flow	> 0.02m^3s^{-1} [> 20ls^{-1}]
Q Selection	Q_{10} to Q_{100}
Classification	Micro to large
Plant Capacity	> 10kW
Grosshead	10m to 1000m
Turbine types	All
Penstock distance	386m to 15km

Table 6.11 Final parameter range restrictions when searching the YDB for viable hydropower configurations

The 20 million plus cells of the YDB raster were searched in parallel on the Eddie Linux Compute Cluster Mark 3 in blocks of 10,000 cells utilising approximately 30 nodes at any one time.

6.10 Cost sensitivity analysis

Whether a scheme is considered viable or not could be down to the parameters set and assumptions made in the algorithm. Therefore, a cost sensitivity analysis was conducted using the final conflict free results at 5% discount rate by reanalysing those configurations with ±10%, 25% and 50% increase/decrease of the original parameter (see section 7.6). Due to time limitations, not every parameter could be tested but the following key variables were: Conversion rate, Surface flow, B factor, F factor, River width, River depth, Gross-head, Penstock Length, Access road length, Transmission line length, Penstock diameter, Head-loss, Energy available, Turbine efficiency, Income per MWh (Energy price), Generator and control cost, Turbine and governor cost, Installation of energy equipment cost, Access road cost, Transmission line cost, Substation and transformer cost, Installation of substation cost, Penstock cost, Installation of penstock cost, Development cost, Feasibility study cost, Miscellaneous costs and Operation and maintenance costs.

6.11 Chapter conclusions

- The hydropower search algorithm developed in this project was based on the RETScreen Small Hydro project model.
- RETScreen has been used widely in hydropower development and tested extensively within the literature.
- Various assumptions were made during the costing of each potential scheme, which could differ from real life. Particularly the fact that scheme construction time was assumed to be instantaneous and penstocks made from GRP could make schemes more attractive than reality. Therefore, it is important that this tool is used as an initial assessment only.
- A large assumption was also made that the RETScreen model is applicable for schemes above 50MW, as the RETScreen model was designed for up to 50MW.

- The hydropower search algorithm was developed in the R programming language, incorporating the datasets selected/developed elsewhere in this thesis (e.g. flow duration curves, DEM, road network etc).
- It was also assumed that all the energy available from a particular scheme was sold at a fixed price, which again could change, which could make schemes more (or less) attractive than in reality.
- Three discount rates were considered 5%, 10% and 15%.
- To reduce the size of the search, input parameter ranges were selected based on the results of a test catchment, eliminating schemes that are most likely to be unviable.
- A unique hydropower conflict algorithm was developed to eliminate conflicting schemes leaving the optimal (based on NPV) configuration of cascading schemes.
- It was found that restricting the maximum penstock length to 10km increased the net maximum profit from a catchment, and if restricted to 13km the net energy generated was maximised. Hence a cascade system of smaller penstock schemes was more efficient than fewer schemes with very long penstocks.
- As the parameters of the search are all subject to error, a cost sensitivity analysis is required (see chapter 7).

7 Synthesis of results, validation and analysis of the resource assessment

7.1 Introduction

This chapter focuses on the output from the hydropower search algorithm, giving an overview of the results and the characteristics of conflict free schemes deemed viable (i.e. an IRR ≥ 5%). As stated in section 3.4.9, an erroneous gross head could inflate the potential value of a hydropower site, and hence the gross head of randomly selected schemes will be compared to the gross head as calculated from Google Earth and a solution sought to correct for gross head error. An analysis of the costs will be presented, comparing costs per installed kW and the levelised cost of electricity (LCOE) to values given elsewhere in the literature, together with component costs as a percentage of the total. A cost sensitivity analysis is necessary to determine how different variables, both assumed and calculated, impact on the overall costs with view to recognising high sensitivity parameters. Power generation from all the viable sites across the catchment as a daily time-series and power generation from example individual sites will be presented along with a summary of plant capacity factors. Finally, two other model variables will be investigated further: Access road length and transmission line costing for schemes with the required voltage above the maximum transmission line voltage (1100kV).

7.2 An overview of initial results

The full search investigated 1,411,299 intake (river) locations across the YDB, producing a database of 311GB of data (approximately 276 million searches). Approximately 60 million of these had an IRR of ≥ 0%, and approximately 24 million (9%) had an IRR ≥ 5% - i.e. those deemed viable providing the discount rate is ≤ 5% (see figure 7.1). Sites with the same intake and turbine location were grouped and the configuration with the best NPV selected, the others rejected, leaving 2,100,609 configurations with an IRR ≥ 5%. Note that this list includes configurations which would conflict with each other and hence they are not all independently

viable. A map showing the locations of this 'best NPV' dataset is shown in figure 7.2, symbolised by the category of installed capacity.

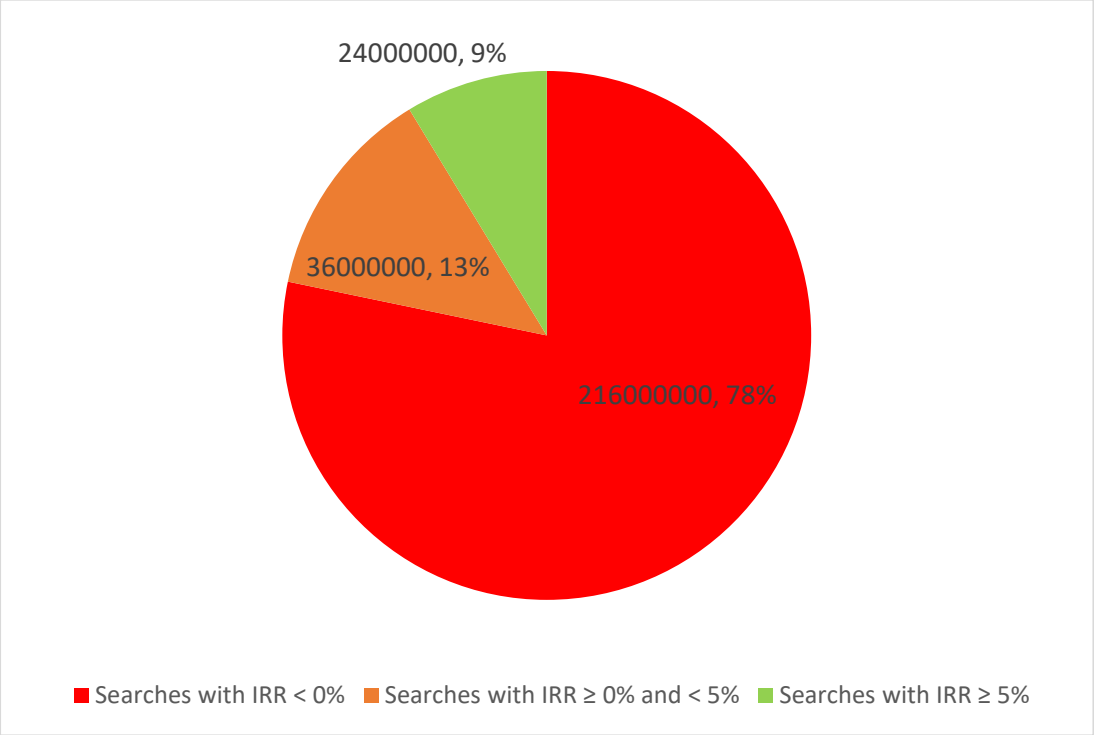


Figure 7.1: Share of IRR range of results from the hydropower search algorithm

The accompanying DVD has the full database of results as a .CSV file (see Appendix 1 for list of files contained on the DVD).

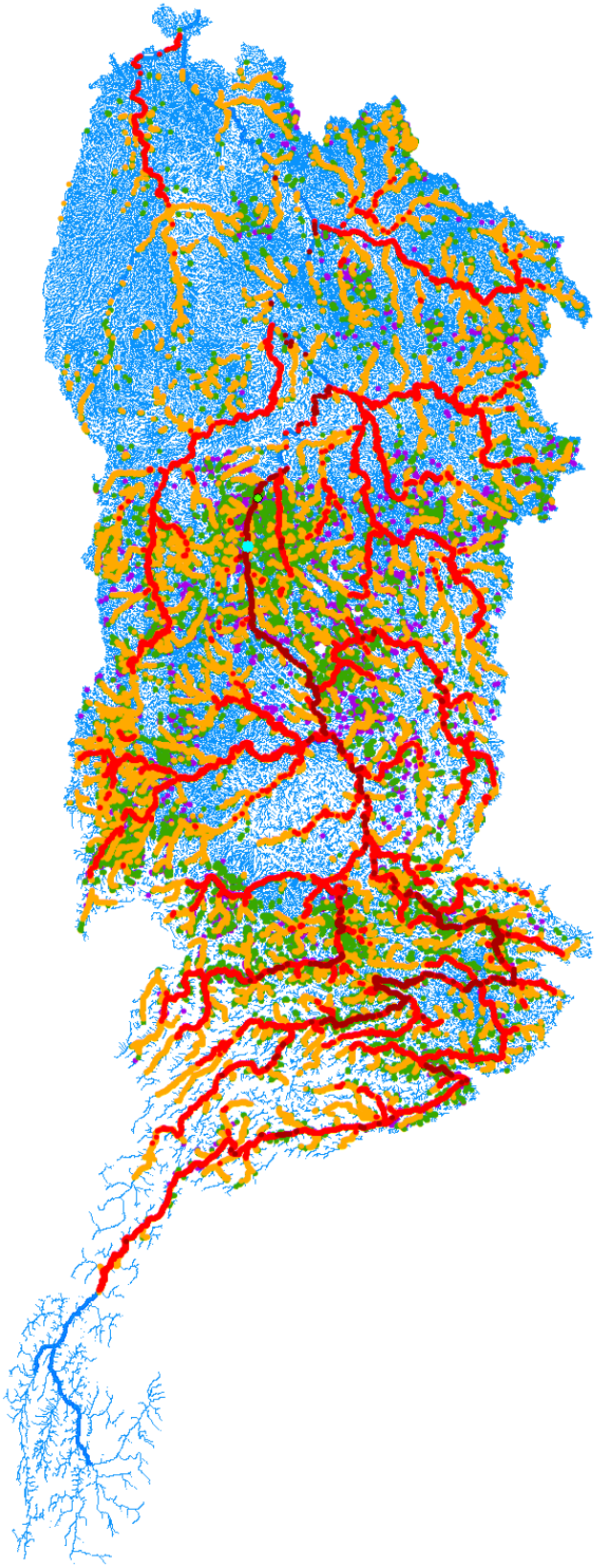


Figure 7.2: Intake locations of viable hydropower sites (where IRR ≥ 5%), symbolised by category of installed capacity category (where dark red = very large ≥ 1GW, red = large ≥ 50MW & < 1GW, orange = small ≥ 1MW & ≤ 50MW, green = mini ≥ 100kW & <1MW, purple = micro ≥10kW & < 100kW, blue = river locations without viable hydropower schemes)

The headings within the database are explained in table 7.1:

Heading	Explanation
IntakePointx	Longitude of intake
Intakepointy	Latitude of intake
DischargePointx	Longitude of turbine
DischargePointy	Latitude of turbine
DesignFlow	Maximum flow of scheme [m^3s^{-1}]
QSelection	Percentile flow of design flow where Q0 is the 100^{th} percentile (highest flow) and Q100 is the minimum flow – i.e. the flow exceeded 100% of the time
Classification	Category of Plant Capacity i.e. micro (<100kW), mini (100kW to 1MW), small (1MW to 50MW), large(50MW to 1GW) or very large (>1GW)
Grosshead	Difference in height between intake and turbine [m]
Turbine	Type of turbine i.e. Francis, Pelton, Turgo, Crossflow, Kaplan or Propeller
NumberofTurbines	Number of turbines (also equals number of penstocks)
PlantCapacityMW	Maximum power generated [MW]
PlantCapacitykW	Maximum power generated [kW]
FeaibilityStudyCost	Cost of feasibility study [RMB]
DevelopmentCost	Cost of development [RMB]
EngineeringCost	Cost of engineering [RMB]
GeneratorandControlCost	Cost of generator and control [RMB]
TurbineandGovernorCost	Cost of turbine and governor [RMB]
InstallationofEnergyEquipmentCost	Cost of installation of generator, control, turbine and governor [RMB]
AccessRoadCost	Cost of making access road [RMB]
TransmissionLineCost	Cost of transmission line and installation [RMB]
SubstationandTransformerCost	Cost of substation and transformer [RMB]
InstallationofSubstationCost	Cost of installation of substation and transformer [RMB]
PenstockCost	Cost of penstock [RMB]
InstallationofPenstockCost	Cost of penstock installation [RMB]
MiscellaneousCost	Other costs, including unforeseen costs and interest during construction [RMB]
TotalInitialCosts	Sum of all costs detailed thus far i.e. installation and construction costs [RMB]
TotalCosts	Sum of all costs plus operation and maintenance costs [RMB]
O and M costs	Operation and maintenance costs, estimated at 3% of the total initial costs over the 25-year life time [RMB]
Total Revenue	Total revenue, calculated by energy generated over 25-year life time [MWh] multiplied by income per MWhr [RMB]
NPV5	Net present value at 5% discount rate [RMB]
NPV10	Net present value at 10% discount rate [RMB]

Heading	Explanation
NPV15	Net present value at 15% discount rate [RMB]
IRR	Internal rate of return [%]
DPP	Discounted payback period at 5% discount rate [years]
CostperkW	Cost per installed kW [RMB]
CostperMWh	Cost per MWh generated – not levelised cost [RMB]
RiverW	Estimated river width [m]
RiverD	Estimated river depth [m]
Intakeheight	Elevation of intake [m]
Turbineheight	Elevation of turbine [m]
FlowPerTurbine	Design flow / number of turbines [m^3s^{-1}]
Turbine Runner Diameter	Diameter of turbine runner of each turbine [m]
Speed	Speed of runner of each turbine [ms^{-1}]
NoJets	Number of jets in Pelton and Turgo type turbines – assumed to be always 1
Turbinepeakefficiency	Maximum efficiency of turbine, 1 = 100%
Peakefficiencyflow	Flow required for maximum efficiency within each turbine [m^3s^{-1}]
MWperunit	Power capacity of each turbine [MW]
EnergyAvailableMWhryr	Mean energy produced per year [MWhr]
PlantCapacityFactor	Power generated compared to power output possible (i.e. Plant Capacity * time) [dimensionless]
UKhomespowered	Potential UK homes powered if all the energy was delivered evenly to an average UK home, based on 4,648kWhr per year
Chinesehomespowered	Potential Chinese homes powered if all the energy was delivered evenly to an average Chinese home, based on 1,349kWhr per year per year
Penstockdist	Length of penstock [m]
Penstockdia	Diameter of penstock [m]
PenstockGrad	Gradient of penstock [%]
PenstockHeadloss	Head-loss within penstock due to friction [m]
AccessRoad	Length of access road [km]
Transmissionlinekm	Length of transmission line [km]
TransmissionLinekV	Voltage of transmission line [kV]
RequiredvoltagekV	Actual voltage required [kV]

Table 7.1: Heading names and explanation within results database on DVD

Only a proportion of these viable results could be implemented as many of the configurations would conflict with one another i.e. the intake or turbine of one configuration is either at the intake or turbine, or between the intake and turbine, of another configuration.

The hydropower conflict algorithm was run three times, setting a minimum of 5%, 10% and then 15% IRR and a summary of the results given in table 7.2, and maps showing intake

locations and installed capacity category depicted in figures 7.3, 7.4 and 7.5. Sites are located across the basin, with a large number located in the highland region (particularly the middle third of the basin) surrounding the low-lying but highly populated Sichuan basin. There is a notable absence of viable stations across the high region to the west on the border of the Tibetan Plateau, due to being drier, cold and at high altitude.

The conflict free hydropower configurations at a minimum IRR of 5%, 10% and 15% discount rate found are also included as a .CSV file (see Appendix 1).

Characteristic	**Minimum IRR 5%**	**Minimum IRR 10%**	**Minimum IRR 15%**
Number of stations	13,604	8,127	4,534
Installed capacity (TW)	1.53	1.44	1.21
Energy available per annum (TWh)	6,351	6,063	5,077
Continuous energy available GW equivalent	725	692	580
Mean Plant Capacity Factor	47.4%	48.1%	47.9%
Number/capacity/% of total energy of very large hydropower stations (≥1GW)	315 (801.7GW) (52.4%)	289 (644.4GW) (44.7%)	223 (504.4GW) (41.7%)
Number/capacity/% of total energy of large hydropower stations (>50MW & <1GW)	2,234 (664.9GW) (43.5%)	2,442 (728.9GW) (50.6%)	2,194 (665.0GW) (55.0%)
Number/capacity/% of total energy of small hydropower stations (≥1MW & ≤50MW)	4,661 (60.6GW) (3.9%)	3,684 (61.3GW) (4.2%)	1,834 (44.5GW) (3.7%)
Number/capacity/% of total energy of mini hydropower stations (≥100kW & <1MW)	4,307 (2.3GW) (0.15%)	1,676 (1.3GW) (0.09%)	283 (251.5MW) (0.25%)
Number/capacity/% of total energy of micro hydropower stations (≥10kW & <100kW)	2,807 (117.4MW) (0.007%)	36 (3.1MW) (0.0002%)	0 (0MW) (0%)

Table 7.2: Summary of viable conflict free stations with minimum IRR of 5%, 10% & 15%

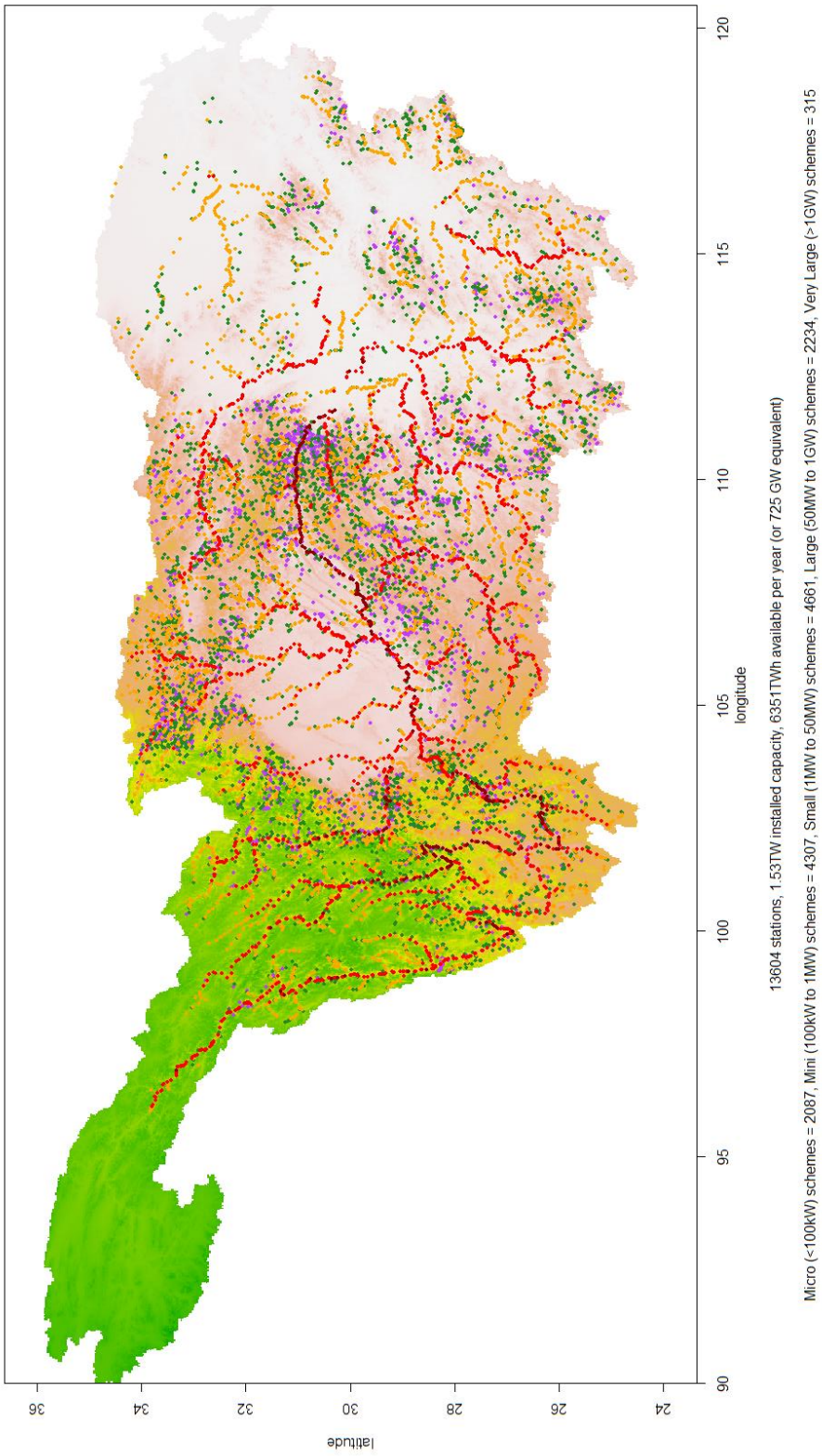


Figure 7.3: Intake locations of conflict free viable hydropower sites (where IRR ≥ 5%), symbolised by category of installed capacity category (where dark red dots = very large ≥ 1GW, red dots = large ≥ 50MW & < 1GW, orange dots = small ≥ 1MW & ≤ 50MW, green dots = mini ≥ 100kW & <1MW, purple dots = micro ≥10kW & < 100kW)

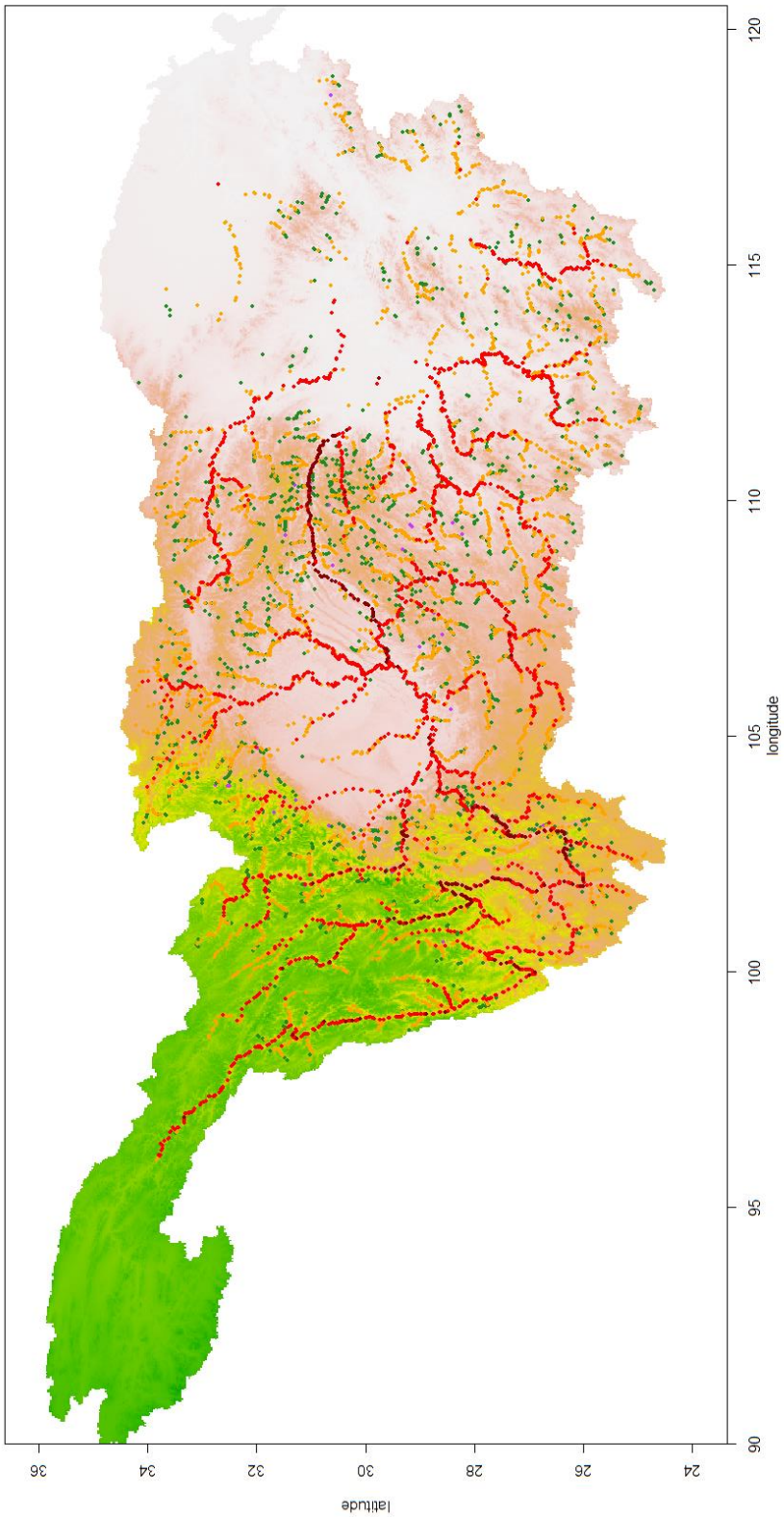


Figure 7.4: Intake locations of conflict free viable hydropower sites (where IRR ≥ 10%), symbolised by category of installed capacity category (where dark red dots= very large ≥ 1GW, red dots = large ≥ 50MW & < 1GW, orange dots = small ≥ 1MW & ≤ 50MW, green dots = mini ≥ 100kW & <1MW, purple dots = micro ≥10kW & < 100kW,)

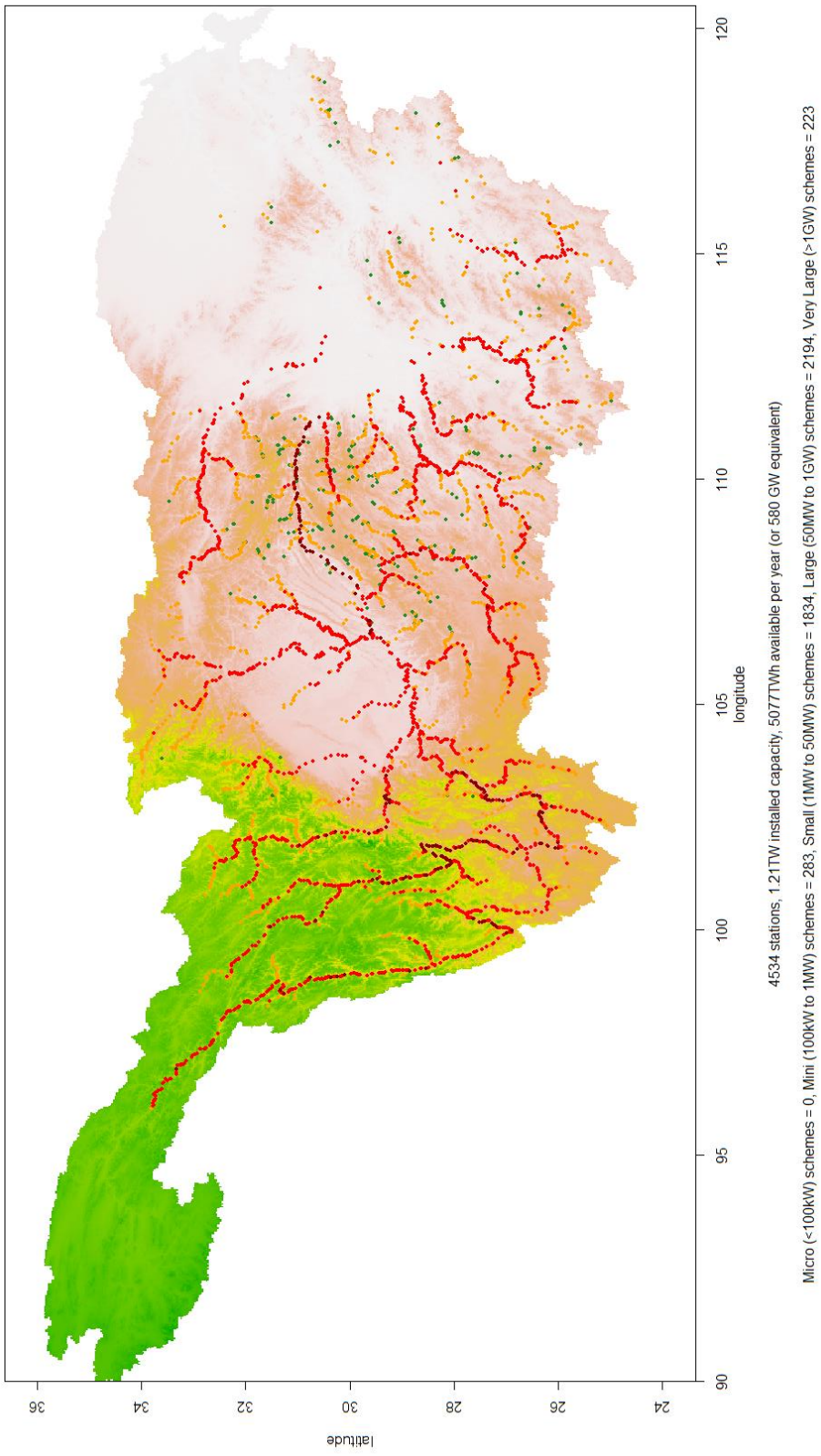


Figure 7.5: Intake locations of conflict free viable hydropower sites (where IRR ≥ 15%), symbolised by category of installed capacity category (where dark red dots= very large ≥ 1GW, red dots = large ≥ 50MW & < 1GW, orange dots = small ≥ 1MW & ≤ 50MW, green dots = mini ≥ 100kW & <1MW, purple dots = micro ≥10kW & < 100kW,)

Although the number of viable stations drops dramatically as the IRR minimum is increased from 5% to 15%, the total installed capacity drops by a relative lesser amount (see figure 7.6) This can be explained due to many of the large and very large power stations having a high IRR, and hence still viable when a high minimum IRR is set, and a large proportion of the installed capacity is within these stations (see figure 7.7, figure 7.8 and figure 7.9 for the case of the minimum of 5%, 10% and 15% IRR, showing both the share of conflict free hydropower stations for each category by number and by installed capacity).

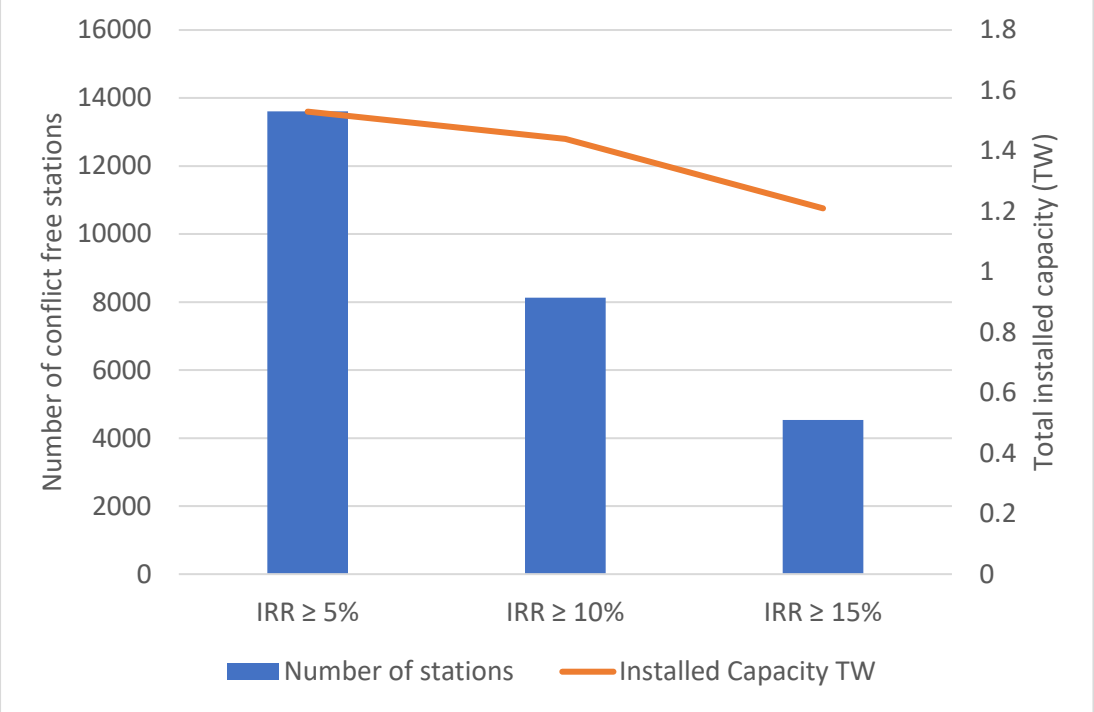


Figure 7.6: Number of power conflict free hydropower stations at IRR 5%, 10% and 15% (blue bars – left axis) and total installed capacity (orange line – right axis)

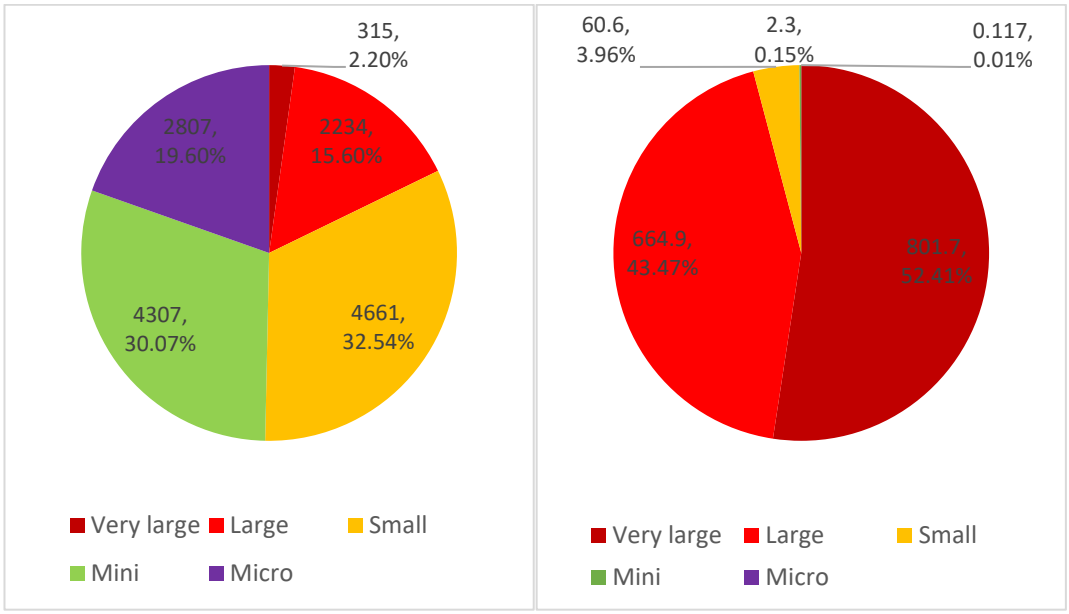


Figure 7.7: Share by category of conflict free hydropower stations with a minimum IRR of 5% by number (left) and by installed capacity (GW) (right)

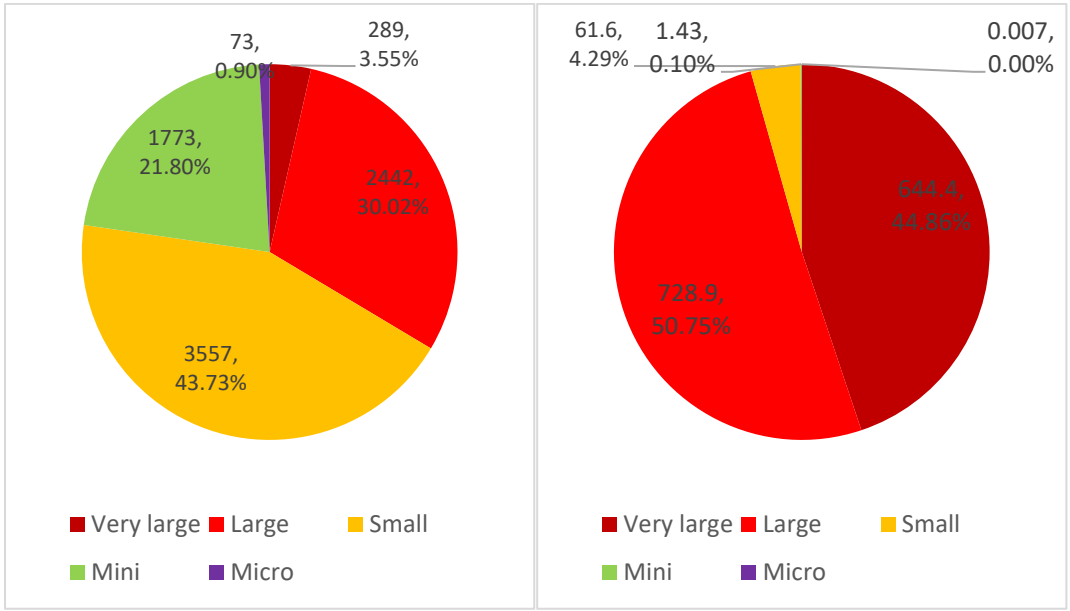


Figure 7.8: Share by category of conflict free hydropower stations with a minimum IRR of 10% by number (left) and by installed capacity (GW) (right)

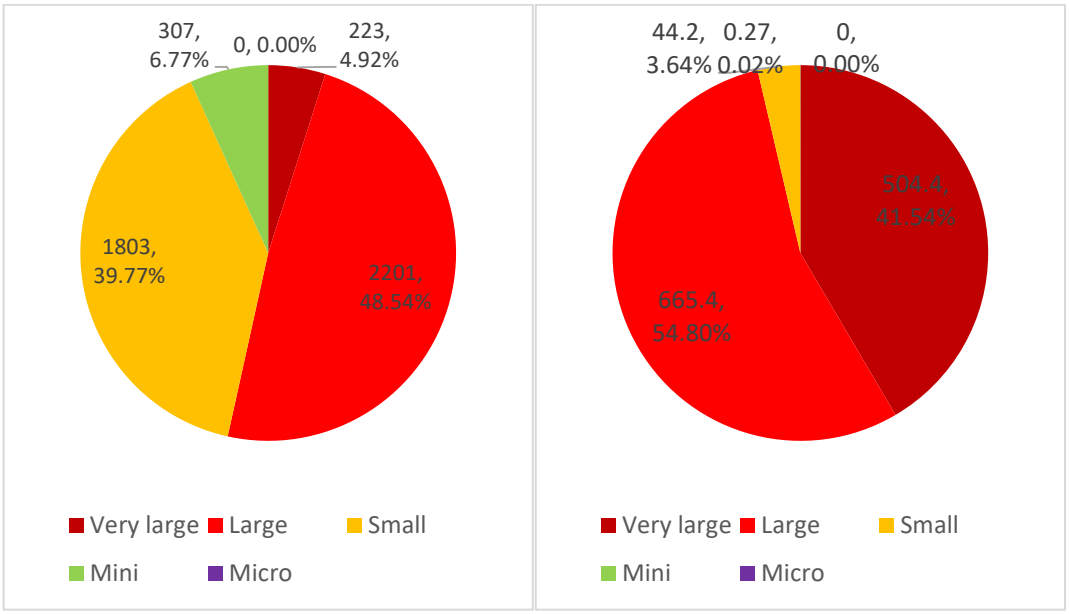


Figure 7.9: Share by category of conflict free hydropower stations with a minimum IRR of 15% by number (left) and by installed capacity (GW) (right)

Although there are too many stations to show here, a more detailed pictorial representation of the stations available can be accessed. An example is shown in figure 7.10.

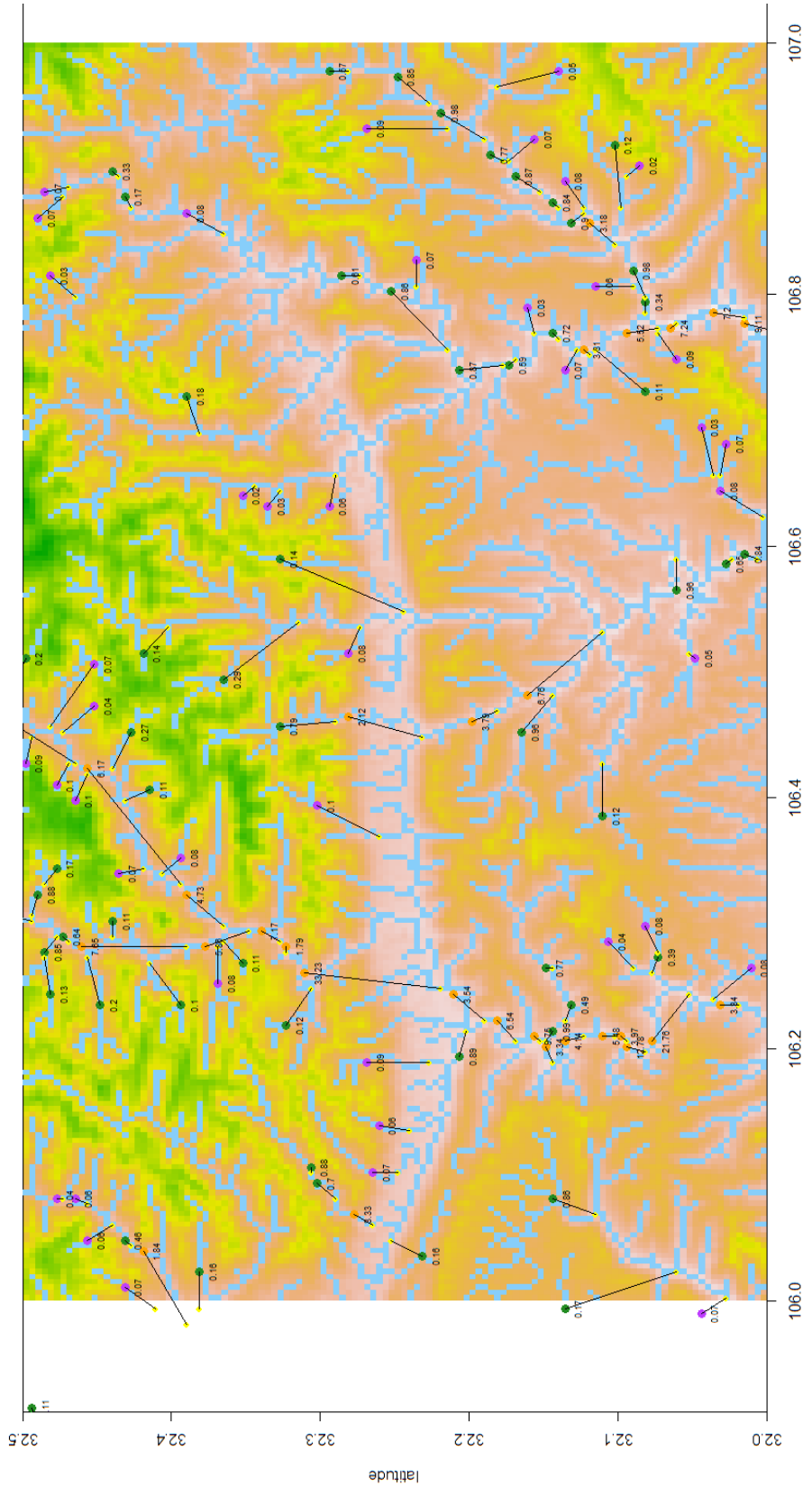


Figure 7.10: Detailed map of viable hydropower stations in the region of 106.0 to 107.0 degrees longitude and 32.0 and 32.5 latitude (yellow dots represent the turbine locations of each viable configuration and the dark red/red/orange/green/purple dots representing the intake location and power category (i.e. very large, large, small, mini or micro). Power capacity is represented by the figure next to the dot in MW and the black line is the penstock between intake and turbine – note the actual penstock may not be a straight line).

7.3 Further investigation of gross head accuracy

Hydropower viability is very much dependent on flow and gross head. The flow data generated by the hydrological model was tested against observed data in chapter 5, and correlates strongly as measured by the objective function results. Here the aim is to check the accuracy of the gross head data by comparing the estimated intake and turbine elevations to those estimated by Google Earth. Google Earth elevation data is by no means perfect and has many errors, but assessments of its accuracy in the literature have been positive, comparing well to STRM (90m) and ASTER (30m) DEMs (Rusli et al., 2014). The methodology to achieve this was as follows:

1. 100 random configurations were selected from the viable list, 20 from each installed power capacity category (i.e. micro, mini, small etc).
2. Using the intake longitude and latitude, the same location visited on Google Earth.
3. The nearest river found visually in Google Earth and the elevation noted (along with the longitude and latitude for later).
4. Steps 3 and 4 repeated for the turbine location.
5. The gross head obtained by the model was compared to the gross head calculated within Google Earth.

The results are presented as a boxplot, shown in figures 7.11 (left box) as an absolute comparison and figure 7.12 (left box) as a comparison as a percentage. Some of the micro-hydro sites were rejected from the study as it was difficult to visually detect the river in Google Earth. A median overestimate of gross head by approximately 40m is significant, potentially making some of the schemes identified unviable, particularly for low head schemes. As this is a 'manual' process it is difficult to test all configurations using this method, but most likely many would reduce in IRR, possibly to a point of becoming non-viable (i.e. IRR < 5%). The 100 locations and results are included on the DVD (see Appendix 1).

Furthermore, the estimated viable energy generation from the original conflict free dataset is greater than 5PWh per annum, even at 15% discount rate, which is significantly above the estimated hydropower potential of all of China (≈2PWh as stated in World Energy Resources: Hydropower (World Energy Council, 2017)).

Reasons for the overestimate are not necessarily due to inaccuracies in the 15-arcseconds DEM dataset used, as this is averaged elevation over the grid square. Rivers tend to flow in steep sided valleys and hence it is quite possible that the intake (and/or turbine) locations estimated elevation is higher than the actual real-world river, which could lead to gross head inaccuracies.

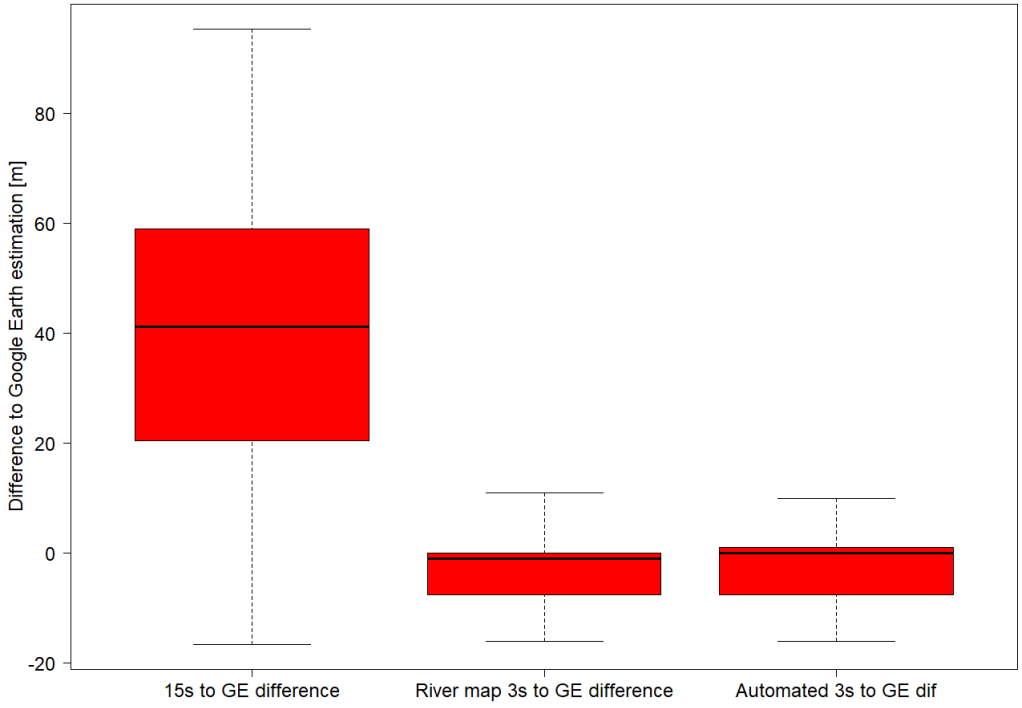


Figure 7.11: Boxplot of absolute difference in gross head of sample schemes estimated by different methods compared to Google Earth. (Left): estimated by 15-arcsecond DEM; (Centre): estimated by finding DEM elevation at nearest point of 3-arcsecond river map to visually identified river on Google Earth; (Right): estimated by automated 3-arcsecond method, finding the lowest DEM cells at same location as 15-arcsecond cell.

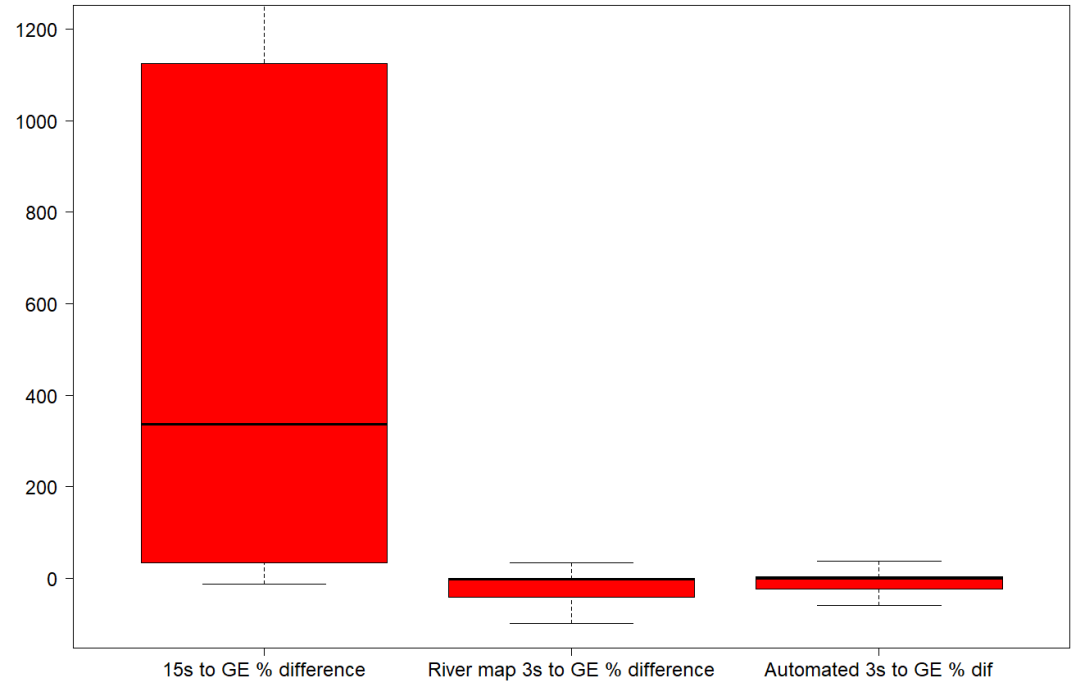


Figure 7.12: Boxplot of percentage difference in gross head of sample schemes estimated by different methods compared to Google Earth. (note: 1m added to all elevations to prevent dividing by zero) (Left): elevations estimated by 15-arcsecond DEM; (Centre): elevations estimated by finding DEM elevation at nearest point of 3-arcsecond river map to visually identified river on Google Earth; (Right): elevations estimated by finding the lowest 3-arcsecond DEM cell at same location as 15-arcsecond cell.

Using a higher resolution dataset within the hydrological model on a catchment the size of the YDB would require a significant increase in computing power. However, to estimate the improvement in accuracy if a higher resolution dataset was used (i.e. 3-arcseconds DEM), the following methodology was employed on the random 100 hydropower configurations from above, utilising the river map derived from the 3-arcseconds flow direction grid:

1. The nearest river point on the river map (derived from the 3-arcseconds flow direction grid) to the river co-ordinates of the intake visually identified from Google Earth (as in step 3 on page 268) was found automatically using the *FNN package* (see section 6.5.1).
2. The height of the DEM was automatically extracted at the river point found in step 1.
3. Steps 1 and 2 were repeated for the turbine location.

4. The gross head of the intake to turbine at the 3-arcseconds DEM resolution was calculated and compared to the gross head estimated from Google Earth.

Comparison of gross head using this method and Google Earth estimations are also shown on the boxplots in figures 7.11 and 7.12 (centre boxes), which give a median value of less than 5m underestimate compared to Google Earth. This demonstrates that more accurate, and conservative, results would be realised if a higher resolution DEM was utilised.

Although the Google Earth DEM is favourably appraised, caution must be issued in its use as it appears that the elevation given depends on the bathymetry of the river i.e. the elevation of the river bed, as the elevation can change rapidly across the width of a river. It is assumed in this hydropower model that the intake and turbine elevations are at the river surface. On several of the river locations inspected in Google Earth, a 'hump' in the river could be seen, where the river rapidly increases in elevation and then decreases – an impossible occurrence in the real-world (see figure 7.13).

Due to the, sometimes dramatic, overestimate of gross head using the 15-arcsecond DEM, it was decided to find the 3-arcsecond DEM estimated gross-head of all the schemes in the original 2,100,609 Best NPV dataset to re-assess the hydropower details of each scheme. The process described above requires visual identification of the river on Google Earth. Hence it is only useful in confirming that employing a higher resolution dataset would lead to an increase in accuracy of estimating the gross head. Therefore, a fully automated methodology was sought and hence the following solution was implemented:

1. The original longitude and latitude of each intake was used to identify and isolate the underlying 15-arcsecond grid square of the original HydroSHEDS DEM raster.
2. The identified grid square from step 1 was increased in resolution to 3-arcssecond (see figure 7.14) using the disaggregate function (i.e. the 15-arcsecond grid split into twenty-five 3-arcsecond grid squares in a 5 x 5 raster) and the values of each set to 1.
3. The disaggregated grid square was multiplied by the HydroSHEDS 3-arcseconds DEM – this isolated the twenty-five 3-arcsecond DEM grid squares at the location of the previously found 15-arcseconds grid square.
4. The minimum elevation of the twenty-five 3-arcsecond DEM grid squares was found and set as the new intake height.
5. Steps 1 to 4 repeated for the turbine location.
6. The new 3-arcsecond gross head calculated.

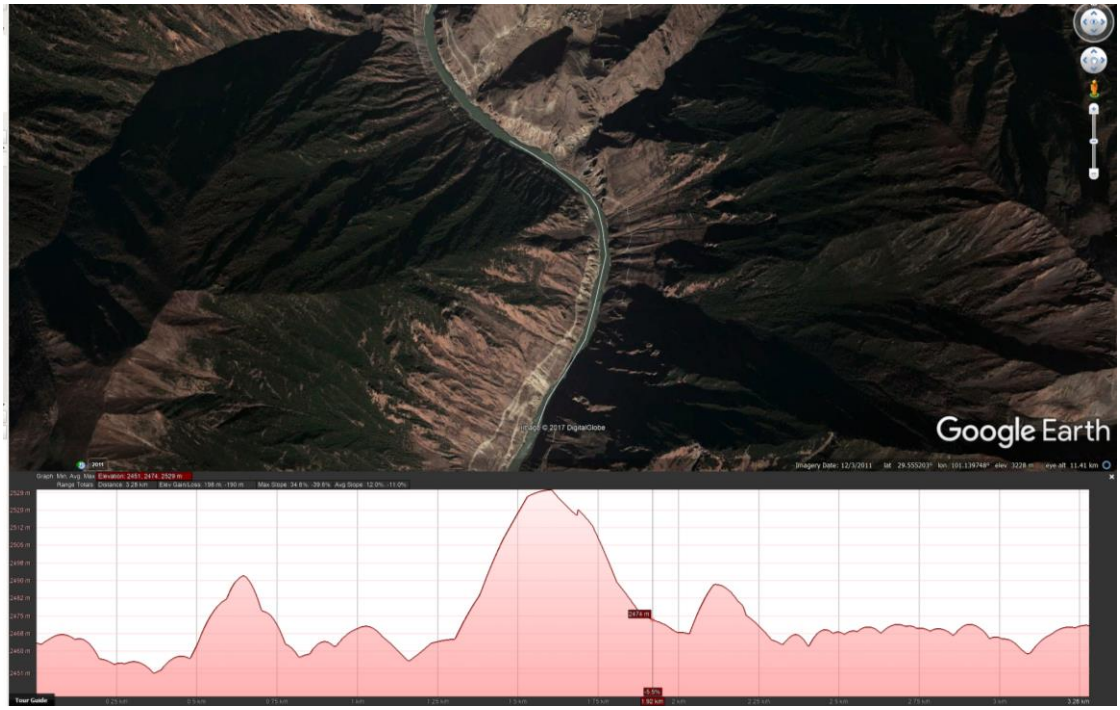


Figure 7.13: An elevation profile of a path drawn along a river at one of the sample hydropower schemes showing an impossible 'hump' in the river, demonstrating imperfections in the Google Earth estimations of gross head.

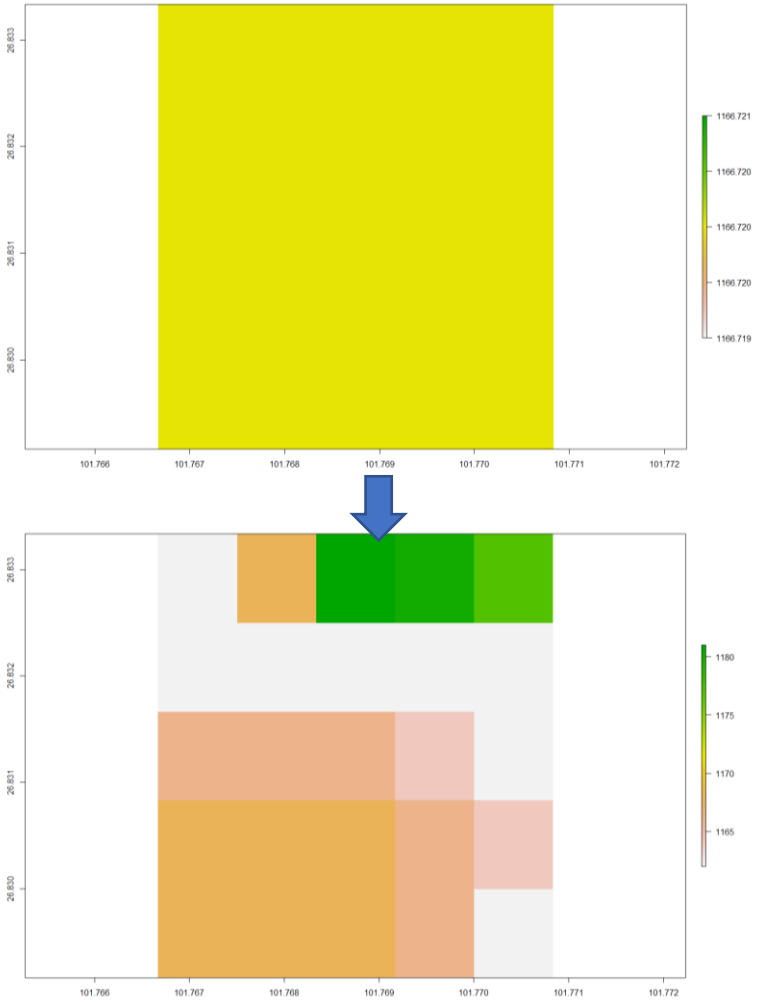


Figure 7.14: Finding 3-arcsecond raster DEM cells at location of original 15-arcsecond DEM and finding the minimum value (in this example the intake elevation was 1172m at 15-arcsecond vs 1162m at 3-arcsecond)

Figures 7.11 and 7.12 (right box), presented earlier, show the results of this methodology compared to the elevation of rivers found within Google Earth when applied to the 100 sample schemes. The median is closer to 0m and 0% than the other methods suggesting a higher accuracy and therefore an acceptable methodology.

Ideally this methodology would have been integrated into the search algorithm at the start of the full search but re-processing of all 276 million searches was impossible in the timeframe allowed. It is likely that a large majority of the unviable searches would still be unviable when integrating the gross head calculated by the 3-arcsecond DEM as in most cases the gross head would be lower, reducing the energy generated. However, surprisingly, some of the schemes

could become more profitable with a lower gross head. As an example, a large flowing river (a mainstream river for example) with an exaggerated high gross head could require multiple Pelton or Turgo turbines (and penstocks etc) to cope with the head/flow design. A lower gross head however would result in replacing the many Pelton turbines with fewer Francis turbines and therefore lowering construction costs, potentially lowering the cost per kW.

Each configuration within the Best NPV dataset (2,100,169 rows of data) was re-assessed with the 3-arcsecond calculated gross head with the following assumptions to reduce the search time:

1. The penstock length and access road length of each configuration was kept the same as the original.
2. The design flow (i.e. Q percentile) was kept the same as the original for each configuration.

However, for each configuration there was still the potential for several new designs as the new head/flow characteristics could enable multiple turbine types, and the power produced may differ from the original requiring a new transmission voltage and/or distance.

This second search produced 4,260,069 new configurations of which 2,151,916 had an IRR ≥ 5%. Those schemes with both identical intake and turbine locations were filtered so only the best performing scheme (by NPV) was selected, leaving 1,346,678 unique intake/turbine location configurations (included on the DVD – see Appendix 1). The hydropower conflict algorithm was once again applied at 5%, 10% and 15% minimum IRR and the results given in table 7.3 (these three sets of results are also included on the DVD – see Appendix 1 for more details).

Characteristic	Minimum IRR 5%	Minimum IRR 10%	Minimum IRR 15%
Number of stations	2825	1509	701
Installed capacity (GW)	130.0	103.2	76.2
Energy available per annum (TWh)	471.4	393.6	311.2
Continuous energy available GW equivalent	53.8	44.9	35.5
Mean Plant Capacity Factor	35%	39%	45%
Number/capacity/% of total energy of very large hydropower stations (≥1GW)	15 (23.5 GW) (18 %)	12 (17.9 GW) (17%)	5 (8.8 GW) (11.5%)
Number/capacity/% of total energy of large hydropower stations (>50MW & <1GW)	417 (88.3 GW) (68%)	339 (71.9 GW) (70%)	251 (59 GW) (78%)
Number/capacity/% of total energy of small hydropower stations ((≥1MW & ≤50MW)	1374 (17.7 GW) (14%)	831 (13.2 GW) (13%)	403 (8.3 GW) (10.9%)
Number/capacity/% of total energy of mini hydropower stations (≥100kW & <1MW)	762 (431.6 MW) (0.33%	293 (184.6 MW) (0.18%)	42 (30 MW) (0.04%)
Number/capacity/% of total energy of micro hydropower stations (≥10kW & <100kW)	257 (15.4 MW) (0.012%)	34 (2.9 MW) (0.002%)	0 (0 MW) (0%)

Table 7.3: Summary of viable conflict free stations with minimum IRR of 5%, 10% & 15% after re-assessing with 3-arcsecond DEM calculated gross head

Maps of optimised (by NPV) conflict free hydropower stations are given in figures 7.15, 7.16 and 7.17 representing the discount rates of 5%, 10% and 15%.

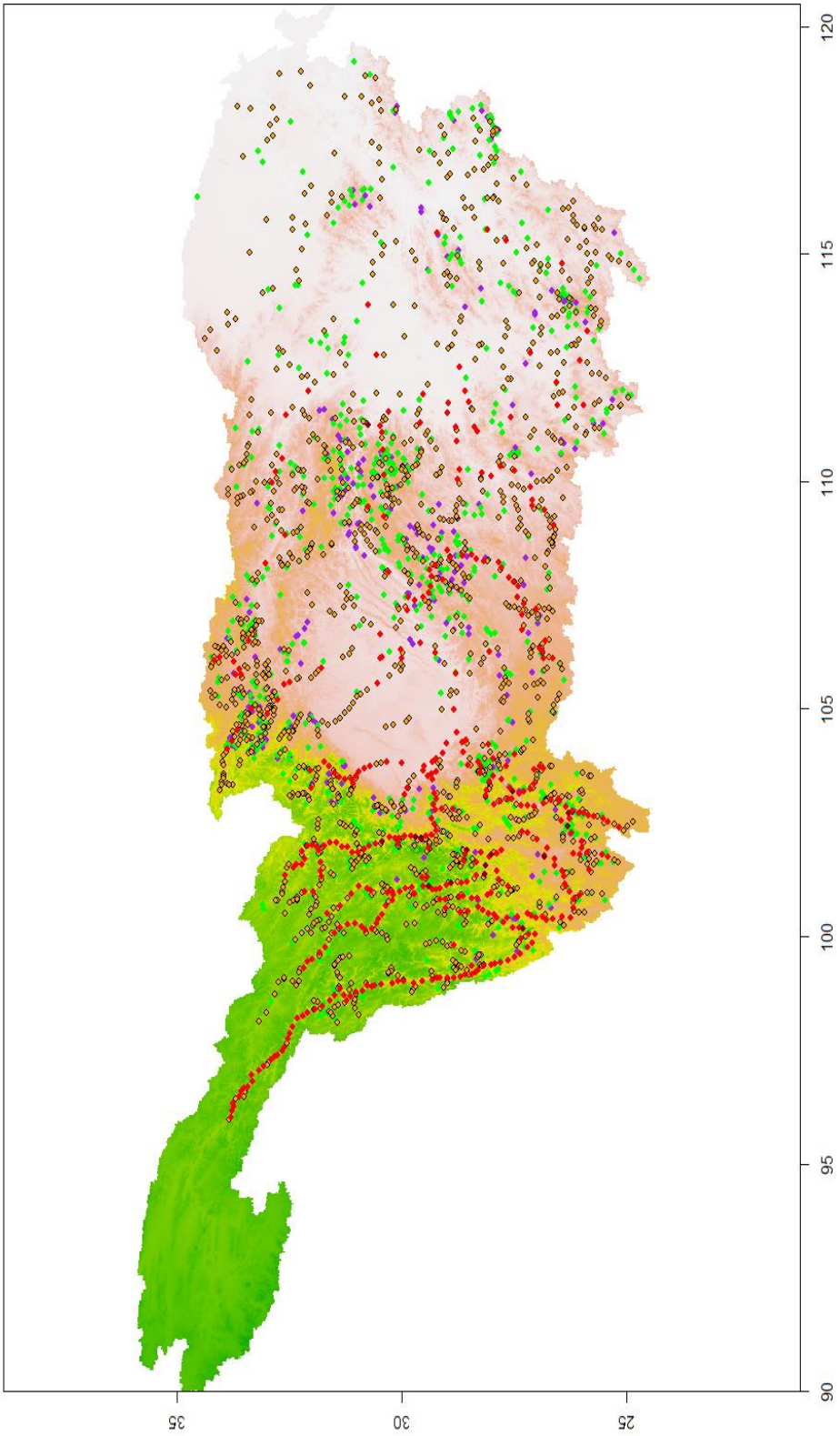


Figure 7.15: Intake locations of conflict free viable hydropower sites (where IRR ≥ 5%) assessed with 3-arcseconds DEM, symbolised by category of installed capacity category (where dark red dots= very large ≥ 1GW, red dots = large ≥ 50MW & < 1GW, orange dots = small ≥ 1MW & ≤ 50MW, green dots = mini ≥ 100kW & <1MW, purple dots = micro ≥10kW & < 100kW,)

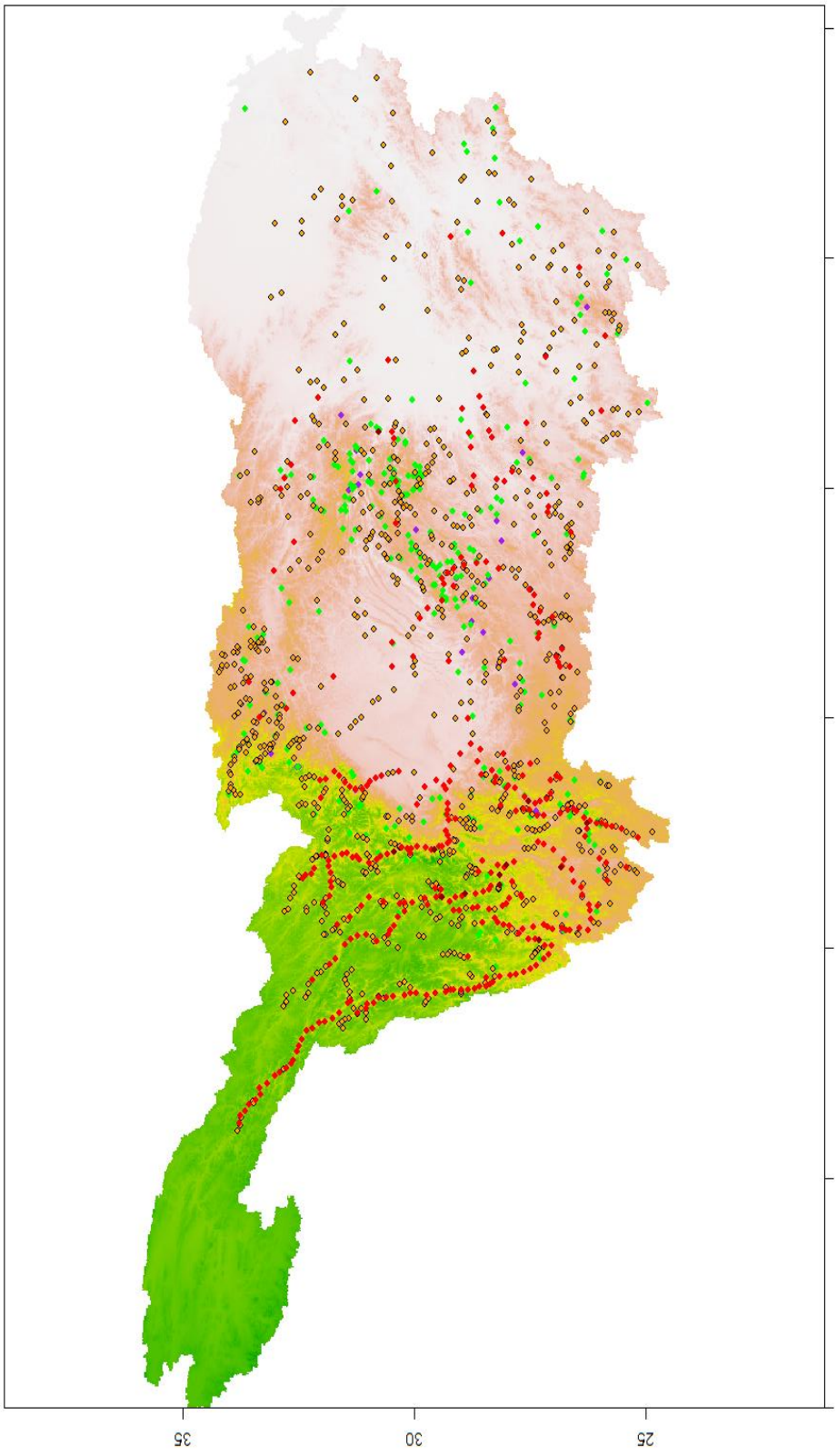


Figure 7.16: Intake locations of conflict free viable hydropower sites (where IRR ≥ 10%) assessed with 3-arcseconds DEM, symbolised by category of installed capacity category (where dark red dots= very large ≥ 1GW, red dots = large ≥ 50MW & < 1GW, orange dots = small ≥ 1MW & ≤ 50MW, green dots = mini ≥ 100kW & <1MW, purple dots = micro ≥10kW & < 100kW,)

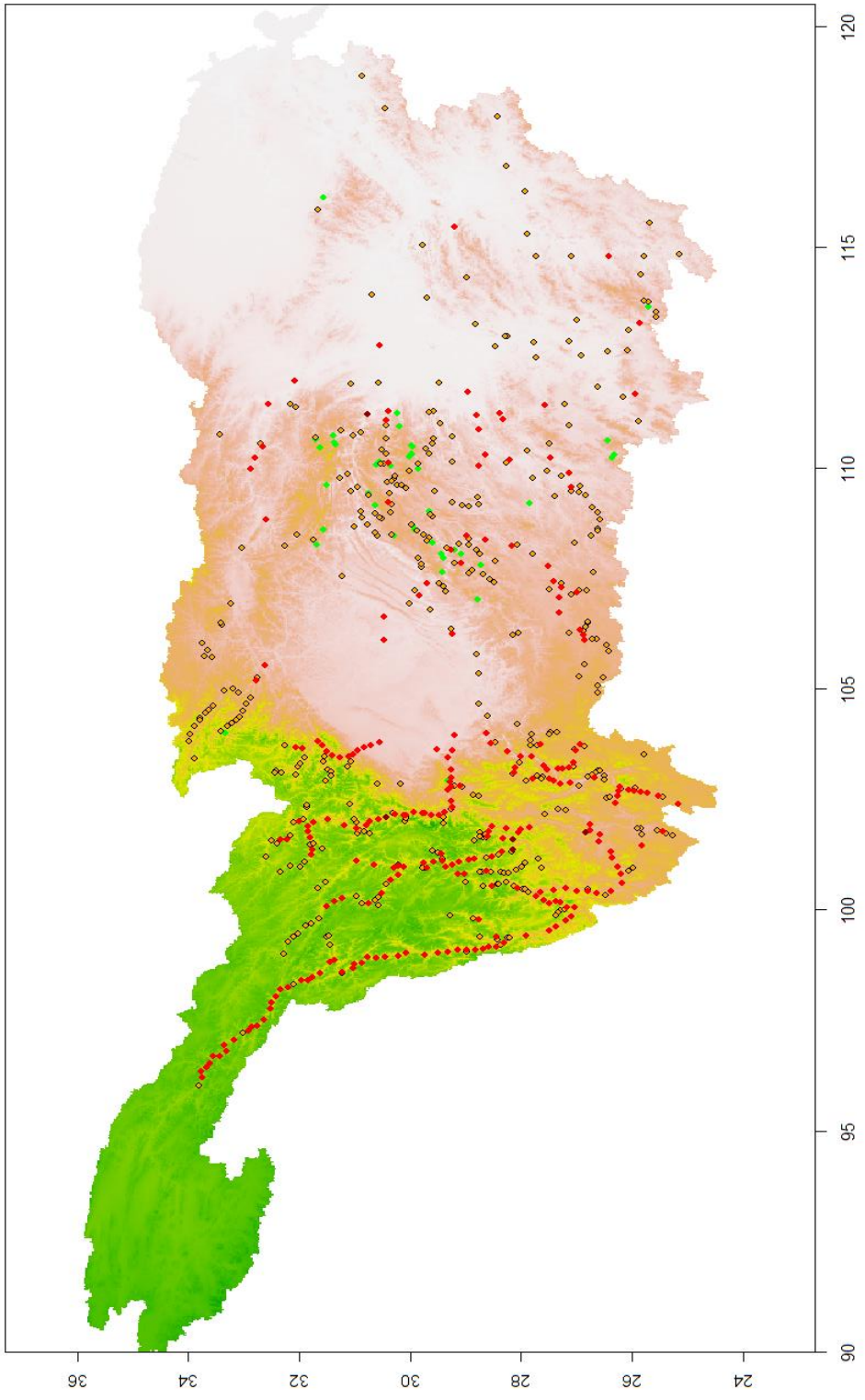


Figure 7.17: Intake locations of conflict free viable hydropower sites (where IRR ≥ 15%) assessed with 3-arcseconds DEM, symbolised by category of installed capacity category (where dark red dots= very large ≥ 1GW, red dots = large ≥ 50MW & < 1GW, orange dots = small ≥ 1MW & ≤ 50MW, green dots = mini ≥ 100kW & <1MW, purple dots = micro ≥10kW & < 100kW,)

Table 7.3 shows a dramatic reduction (approximately a factor of 5) in the number of viable stations at each of the three discount rates compared to when assessed with the 15-arcseconds estimated gross-head, and a corresponding reduction in installed capacity. Most of the larger stations are located west of the Sichuan basin and also a high number of smaller stations located around the Sichuan basin. A large proportion of the installed capacity and energy deliverable is still formed from the 'large' stations category but with a much smaller share of 'very large' stations as shown in figures 7.18, 7.19 and 7.20 representing the discount rates 5%, 10% and 15% respectively.

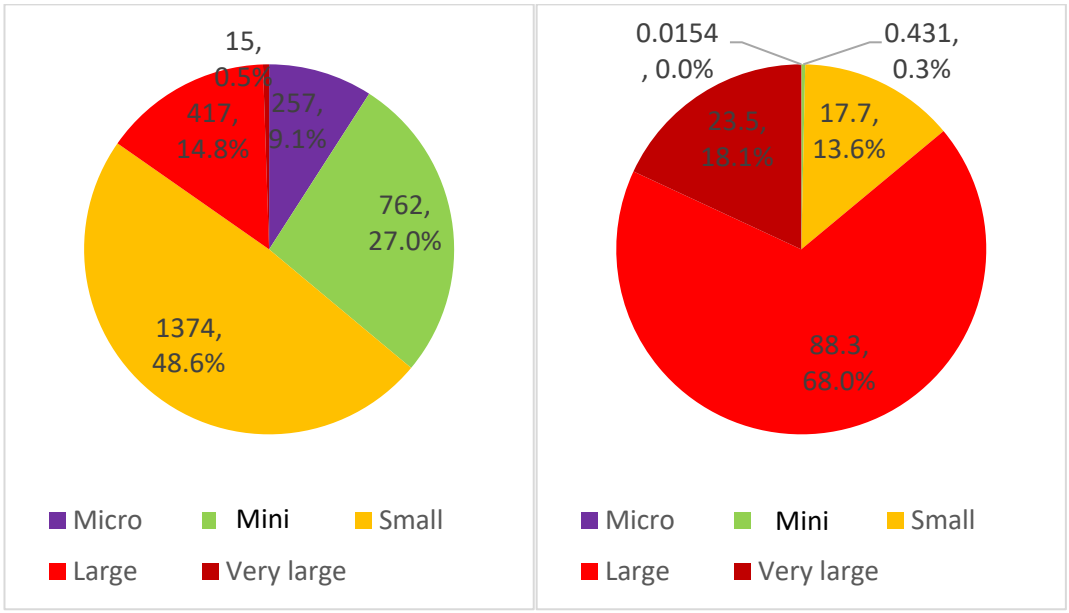


Figure 7.18: Share by category of conflict free hydropower stations when assessed with 3-arcseconds DEM gross-head with a minimum IRR of 5% by number (left) and by installed capacity (GW) (right)

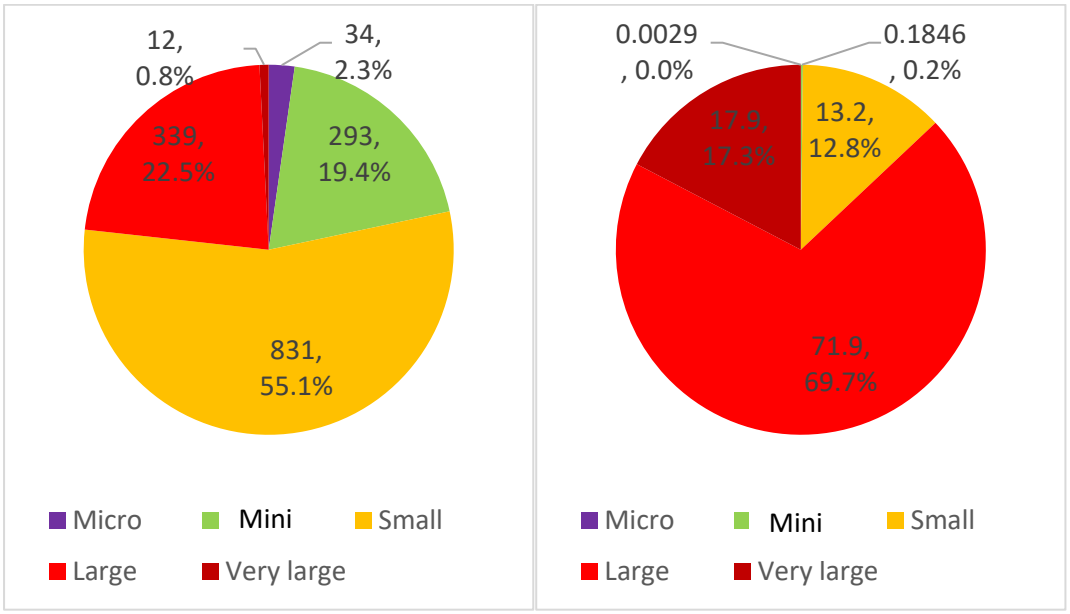


Figure 7.19: Share by category of conflict free hydropower stations when assessed with 3-arcseconds DEM gross-head with a minimum IRR of 10% by number (left) and by installed capacity (GW) (right)

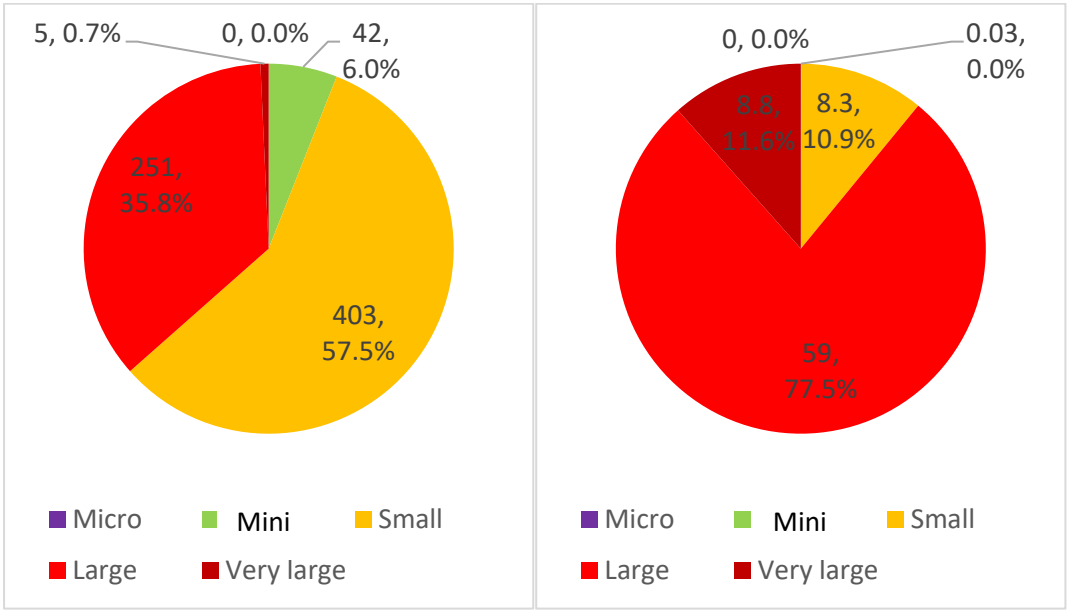


Figure 7.20: Share by category of conflict free hydropower stations when assessed with 3-arcseconds DEM gross-head with a minimum IRR of 15% by number (left) and by installed capacity (GW) (right)

7.4 Characteristics of the conflict free hydropower stations

The design flow and gross head combination of the hydropower configuration is a key component of viability, and a scatter plot of gross head vs design flow of all the stations at 5% IRR is shown in figure 7.21 (note: from now onwards all the analysis will be conducted on the conflict free set assessed with the 3-arcseconds estimated gross-head).

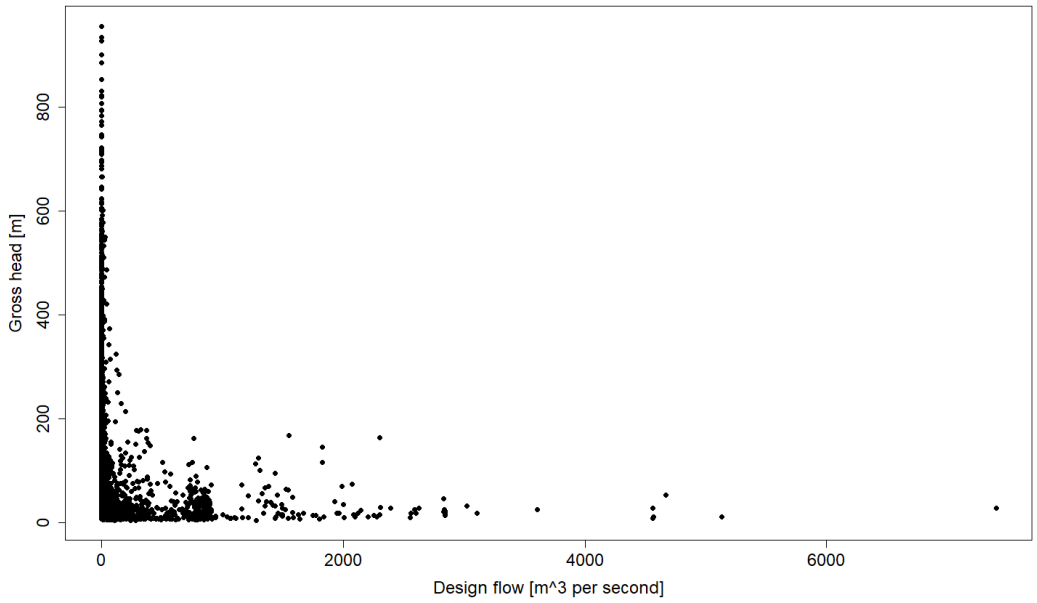


Figure 7.21 Gross head (m) vs design flow (m^3s^{-1}) of all viable stations [IRR≥5%]

A large majority of the viable sites have a design flow less than $100m^3s^{-1}$ (79%), principally because many the YDB rivers are small rivers. Plotting a histogram of design flow of those sites between 0 to $100m^3s^{-1}$ (see figure 7.22) shows that many of the viable configurations (49%) have a flow of less than $10m^3s^{-1}$ and 24% < $1m^3s^{-1}$. Replotting the data with a reduced upper design flow limit of $10m^3s^{-1}$ shows a clear curve of lowering gross head as design flow increases (with outliers) (see figure 7.23).

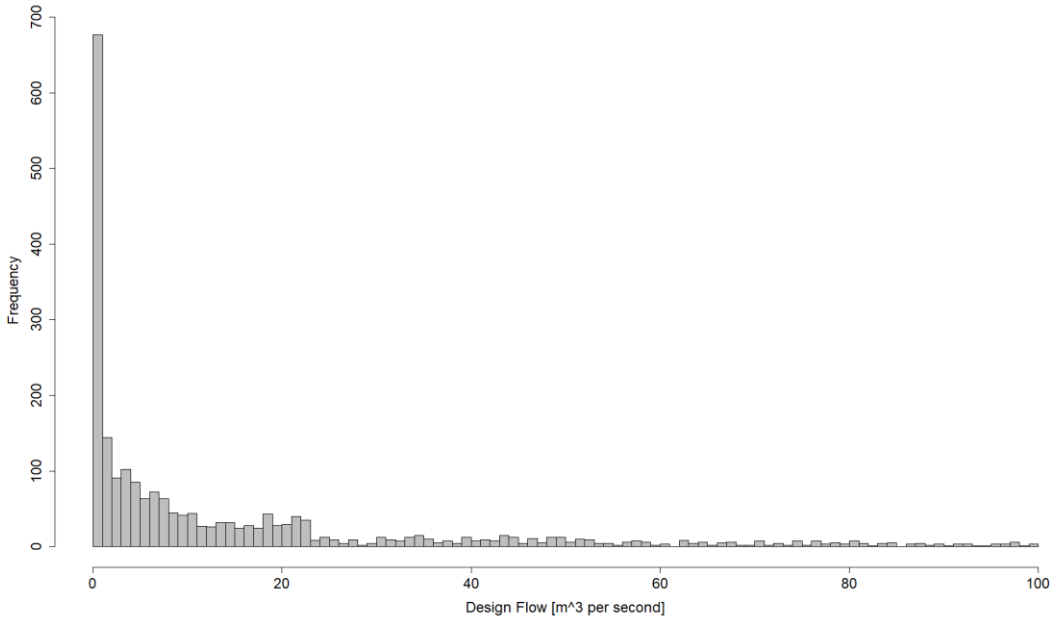


Figure 7.22: Frequency histogram of viable stations (IRR≥5%) by design flow [m^3s^{-1}] when design flow <= 100m^3s^{-1}

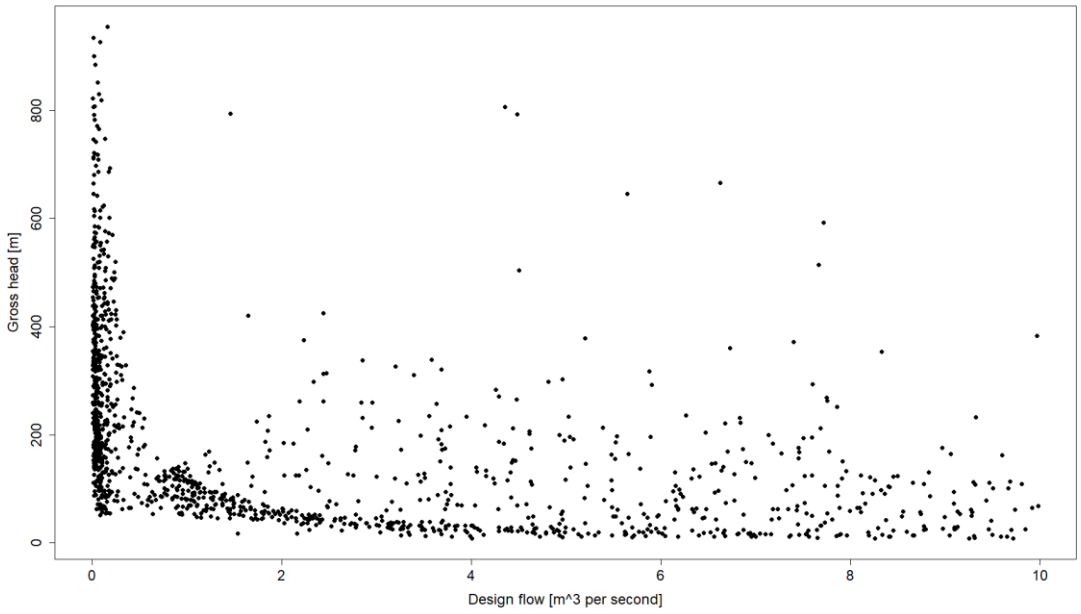


Figure 7.23: Gross head [m] vs design flow [m^3s^{-1}] of viable stations (IRR≥5%) with design flow ≤ 10m^3s^{-1}

Plotting a histogram of gross head at IRR ≥ 5% (figure 7.24) shows a large majority of sites (81.0%) have a gross head < 200m and 45% have a gross head < 50m.

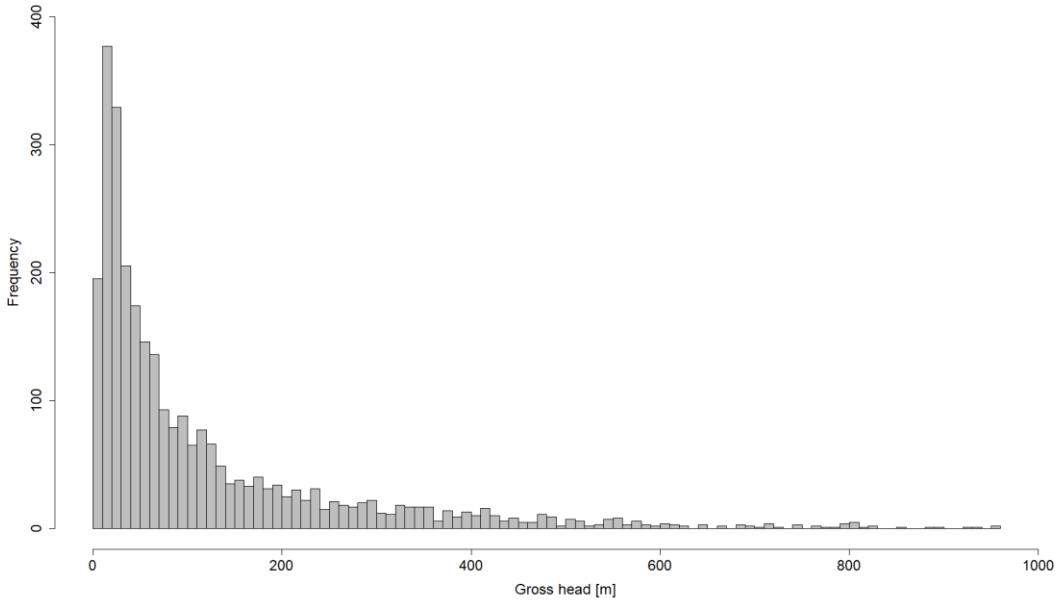


Figure 7.24: Frequency histogram of viable stations (IRR≥5%) by gross head [m]

Scatterplots of gross head vs design flow are shown for viable stations with minimum IRR of 10% and 15% in figures 7.25 and 7.26 respectively.

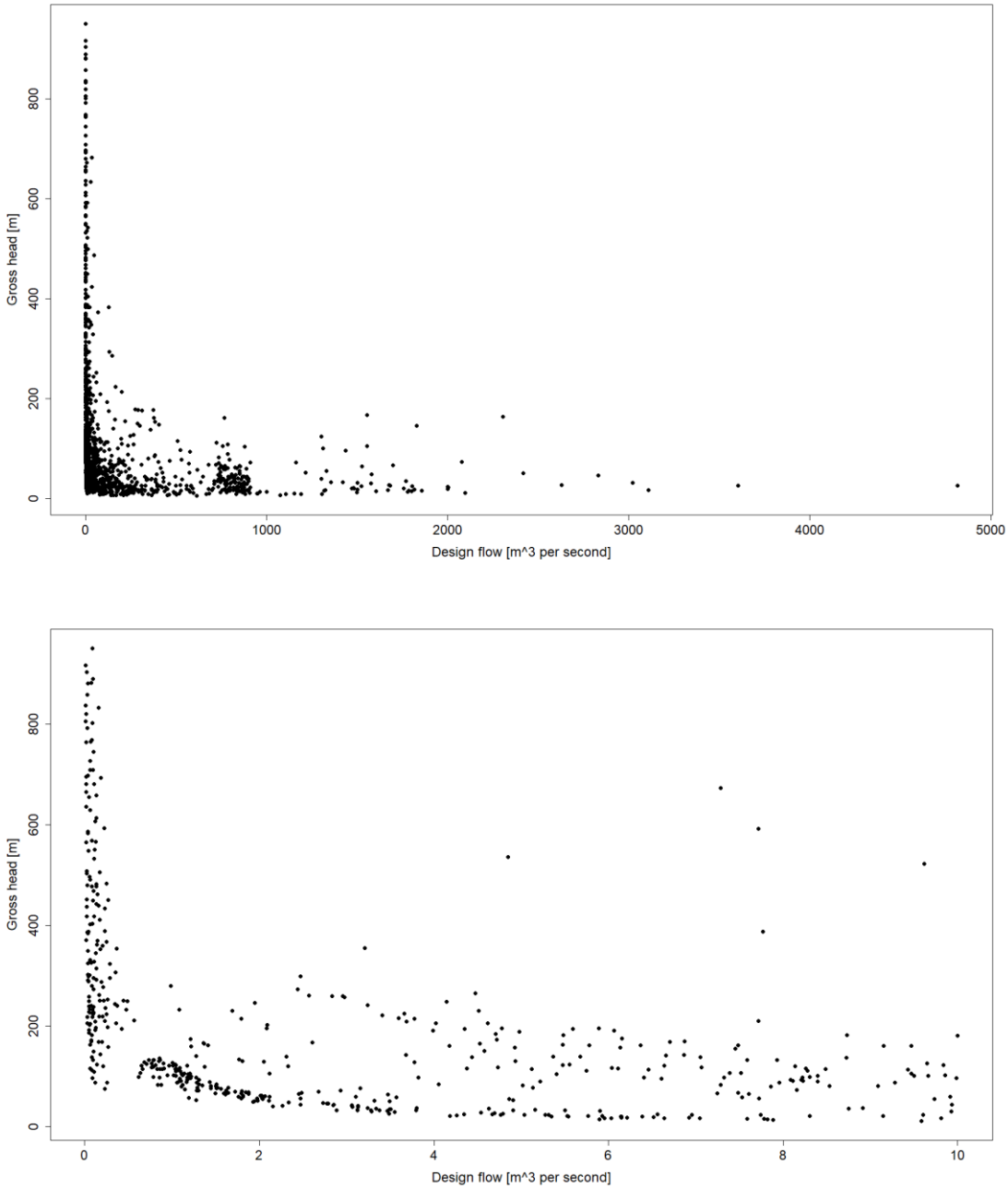


Figure 7.25: Gross head [m] vs design flow [m^3s^{-1}] of all viable stations (IRR≥10%) (top) and those with design flow ≤ 10m^3s^{-1} (bottom)

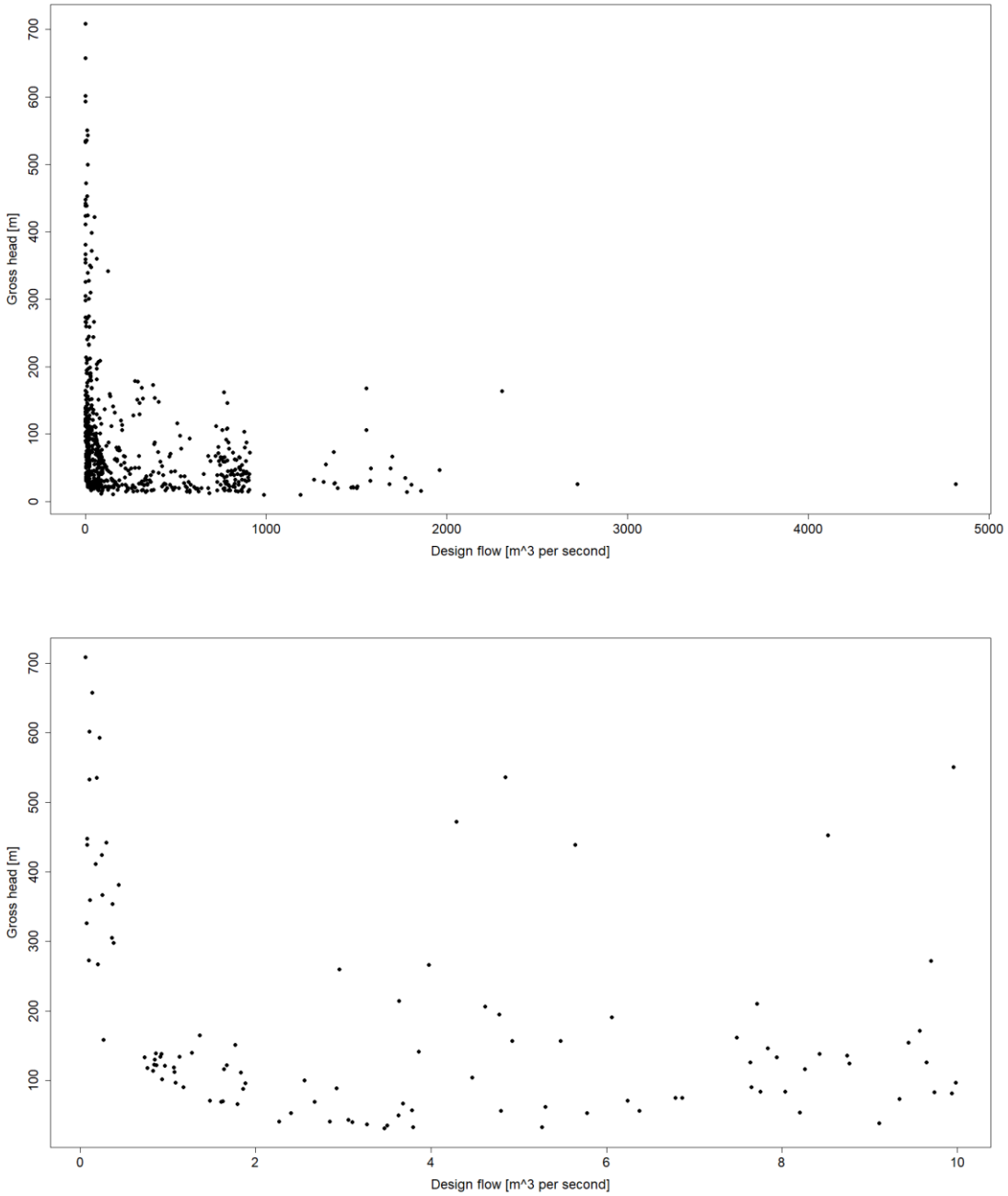


Figure 7.26: Gross head [m] vs design flow [m^3s^{-1}] of all viable stations (IRR≥15%) (top) and those with design flow ≤ 10m^3s^{-1} (bottom)

For all three minimum IRR scenarios, the design flow Q selection was predominantly between Q10 and Q50, with a large percentage between Q15 and Q25 (63%, 60% and 43% for the minimum IRR of 5%, 10% and 15% respectively, represented in figure 7.27).

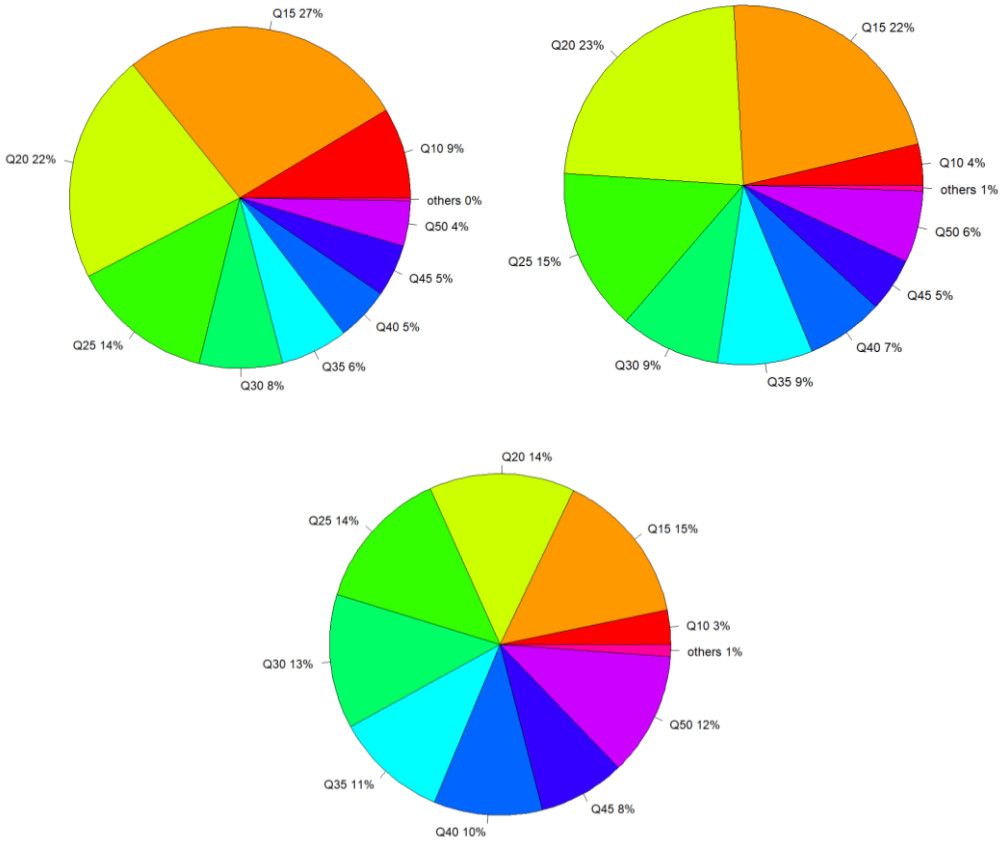


Figure 7.27: Share of Q selection for design flow choice for IRR ≥ 5% (top left), IRR ≥ 10% (top right) and IRR ≥ 15% bottom

Although the search found viable schemes for all six turbine types (Francis, Pelton, Turgo, Kaplan, Propeller and Crossflow), following the conflict algorithm only Francis, Pelton and Propeller configurations were selected. The percentage share of these turbine types is shown in figures 7.28 representing the minimum IRR scenarios of 5%, 10% and 15%.

The number of turbines per configuration varies from 1 to 9, although the majority have 1. Estimated river widths vary from 1.3m to 1353m and most schemes have a river width less than 100m (see figure 7.29). Intake heights of schemes range from 12m up to nearly 4000m, with over 52% of schemes less than 1000m elevation and 34% of schemes below 500m elevation (see figure 7.30).

Penstock diameters range from 0.1m to almost 19m, although a significant proportion of the minimum IRR of 5% results (25%) have a diameter of less than 1m (see figure 7.31). Penstock lengths vary from 400m to 15km and 62% have a length less than 5km (see figure 7.32).

Transmission line lengths vary from 0.07km to 115km, but 80% have a transmission line length of less than 20km (see figure 7.34). Configurations requiring all the available voltage ratings are present with over 42% at 11kV (see figure 7.34). Access road lengths vary from 0km to 18km, but 95% have a length less than 100m and 78% a length less than 10m (see figure 7.35). This is a surprise and will be investigated in the section 7.9 'Further investigation of access road length'.

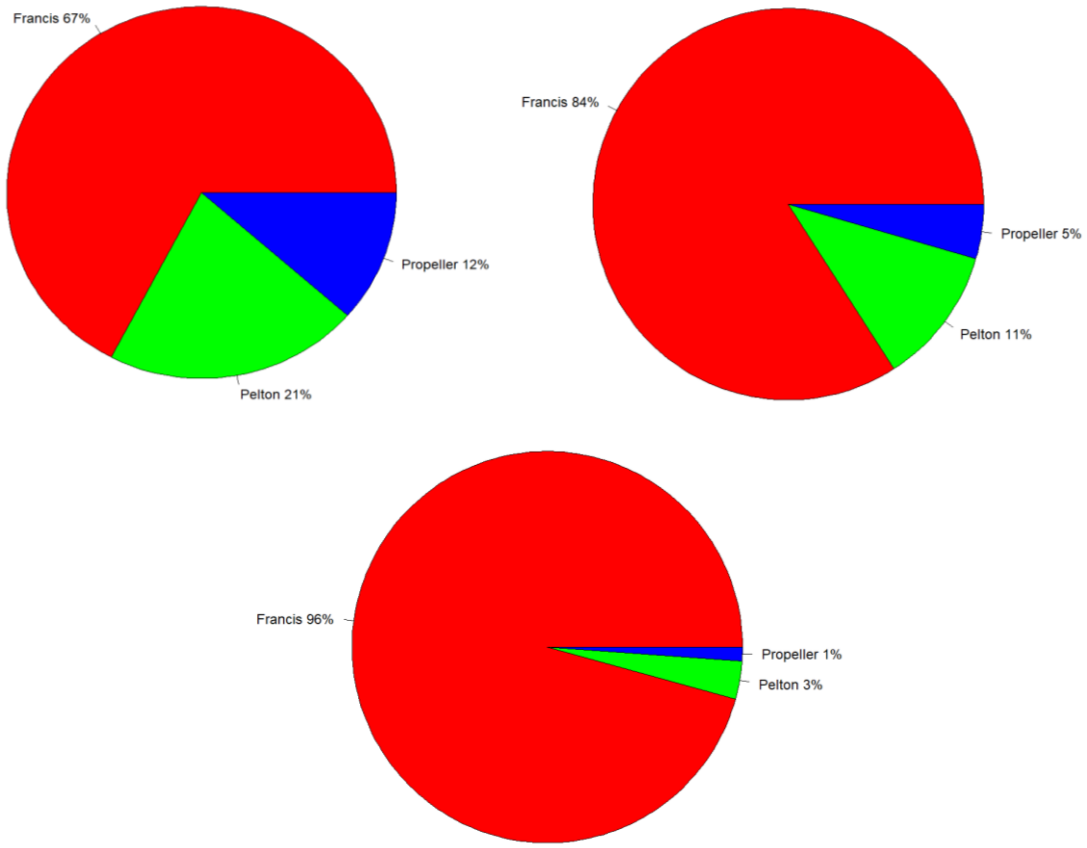


Figure 7.28: Share of turbine type within conflict free configurations for IRR ≥ 5% (top left), IRR ≥ 10% (top right) and IRR ≥ 15%

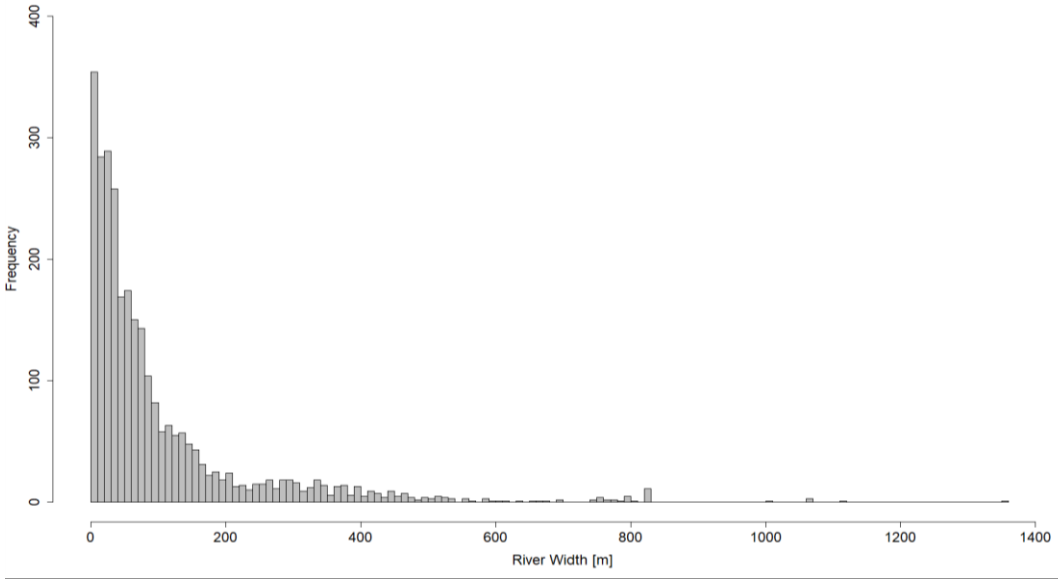


Figure 7.29: Frequency histogram of viable stations (IRR≥5%) by river width [m]

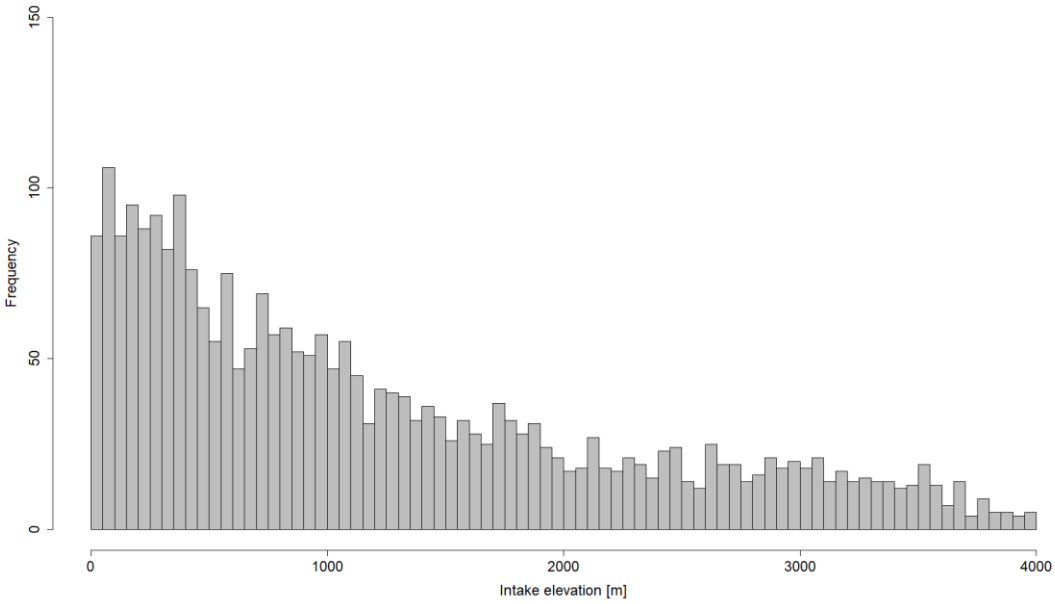


Figure 7.30: Frequency histogram of viable stations (IRR≥5%) by intake height [m]

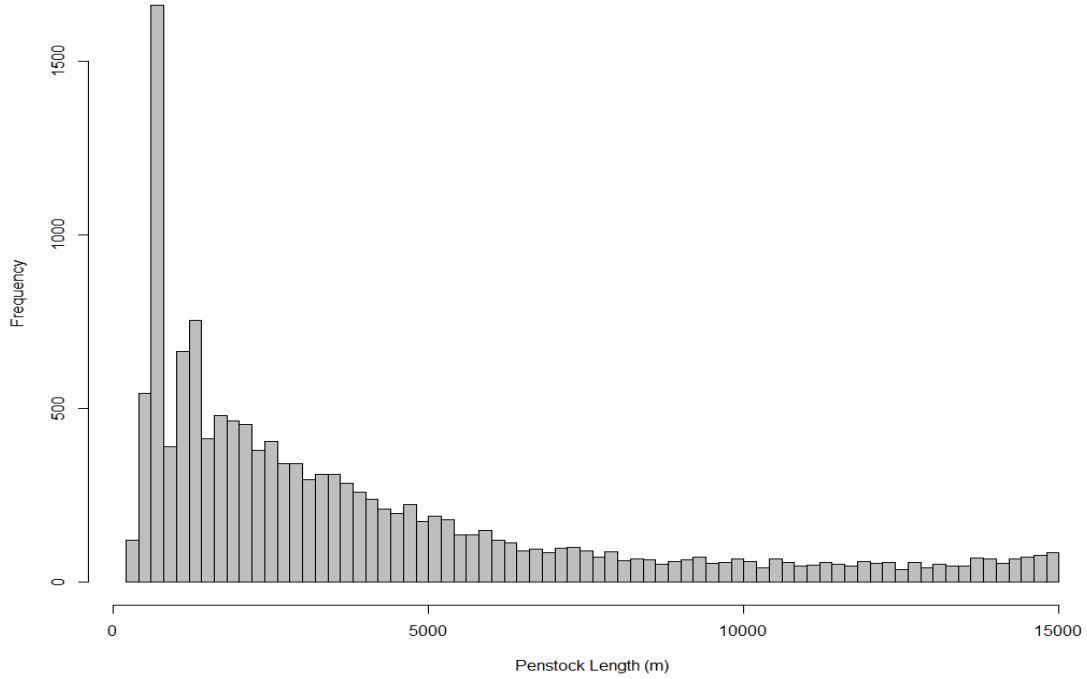


Figure 7.31: Frequency histogram of viable stations (IRR≥5%) by penstock diameter [m]

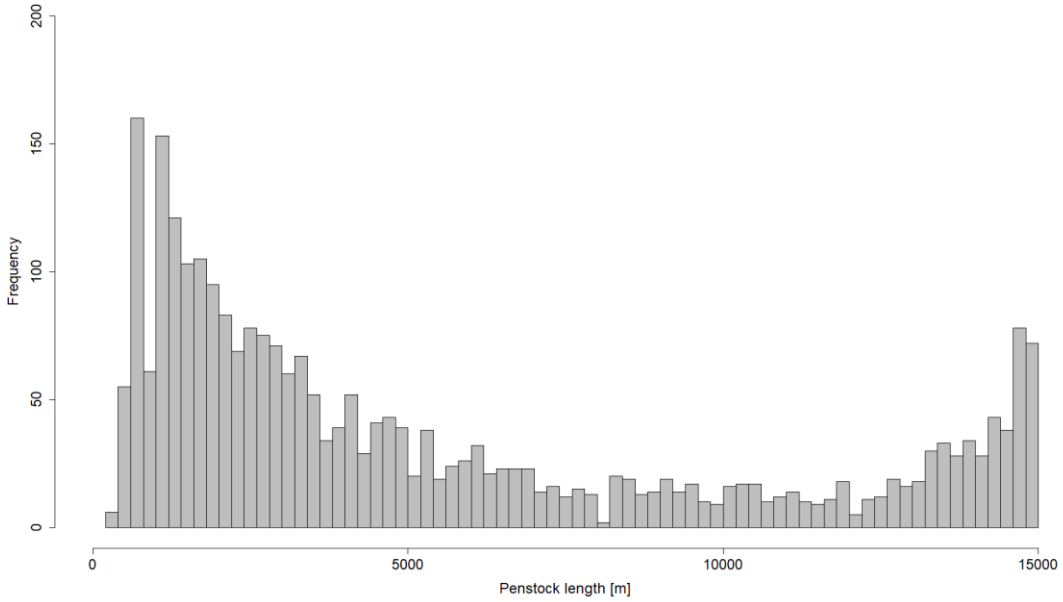


Figure 7.32: Frequency histogram of viable stations (IRR≥5%) by penstock length [m]

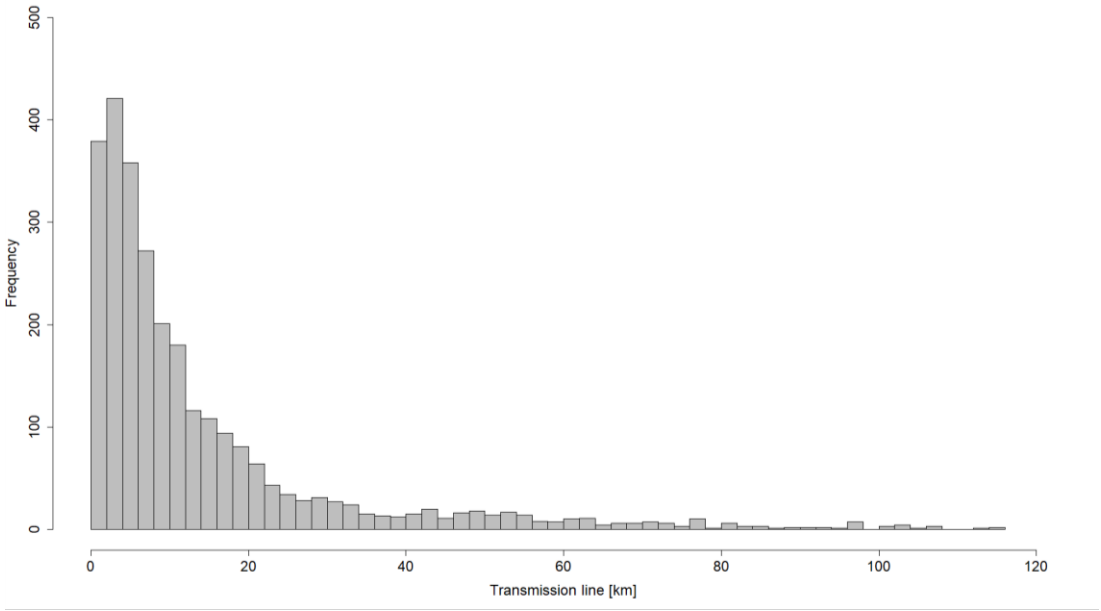


Figure 7.33: Frequency histogram of viable stations (IRR≥5%) by transmission line length [km]

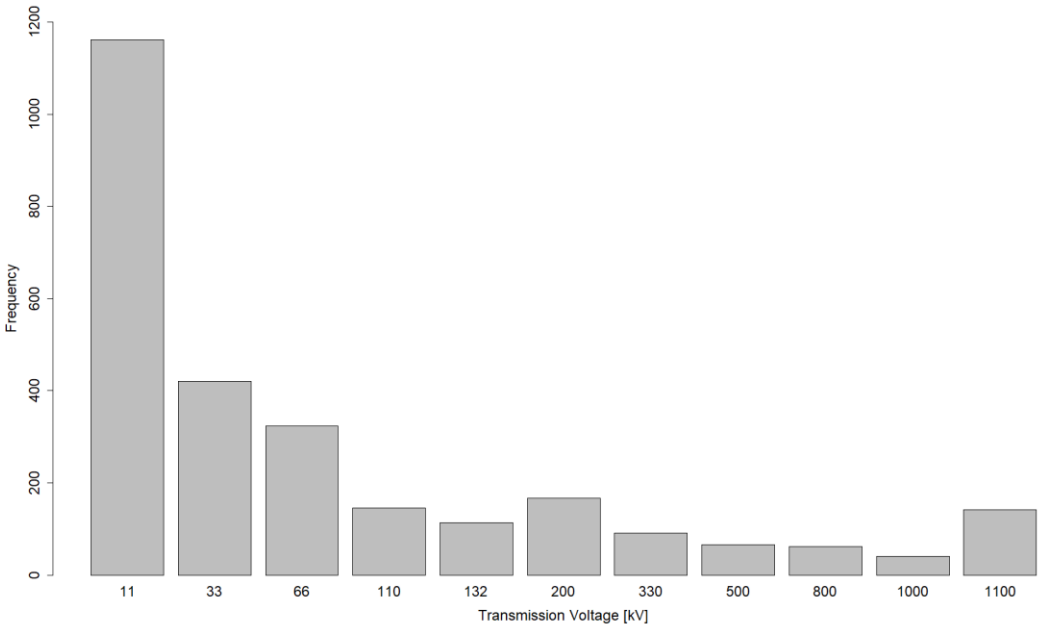


Figure 7.34: Frequency histogram of viable stations (IRR≥5%) by transmission line voltage [kV]

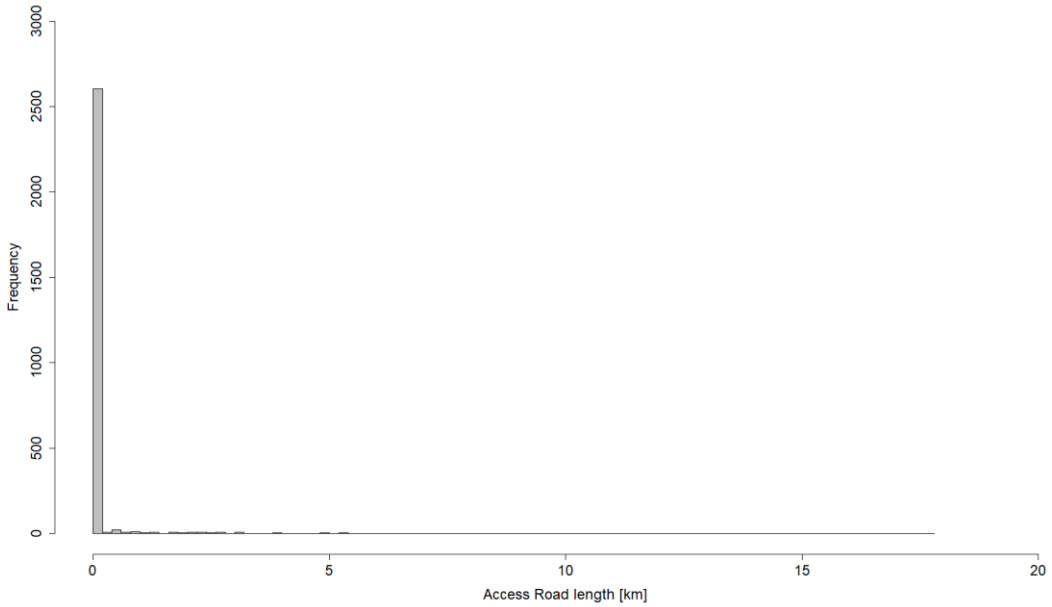


Figure 7.35: Frequency histogram of viable stations (IRR≥5%) by access road length [km]

7.5 Analysis of costs

Costs per installed kW of all the projects at minimum IRR of 5% are shown on the box-and-whisker plot in figure 7.36, where the bottom and top of the boxes are the 25th and 75th percentile respectively (i.e. lower and upper quartile). The whiskers represent:

$$\boldsymbol{Upper\ whisker} = \mathbf{min}(\mathbf{max}(\boldsymbol{Costs})\ \boldsymbol{or}\ \boldsymbol{Q3} + \mathbf{1.5} \times \boldsymbol{IQR}) \qquad (7.1)$$

$$\boldsymbol{Lower\ whisker} = \mathbf{max}(\mathbf{min}(\boldsymbol{Costs})\ \boldsymbol{or}\ \boldsymbol{Q1} - \mathbf{1.5} \times \boldsymbol{IQR}) \qquad (7.2)$$

where Q1 is the lower quartile value (25th percentile), Q3 is the upper quartile value (75th percentile) and IQR is the inter quartile range (i.e. Q1 – Q3). Costs per installed kW generally decrease as projects increase in size:

Installed costs from many published studies are available from within the International Renewable Energy Agency Cost Analysis Series: Hydropower (IRENA, 2012) and shown in figure 7.37. These are in the region of 1000 to 6000 US$ per installed kW, with an average value of between 1000 to 3500 U$ per install kW. This compares well with the estimated costs of the viable configurations as given in figure 7.36.

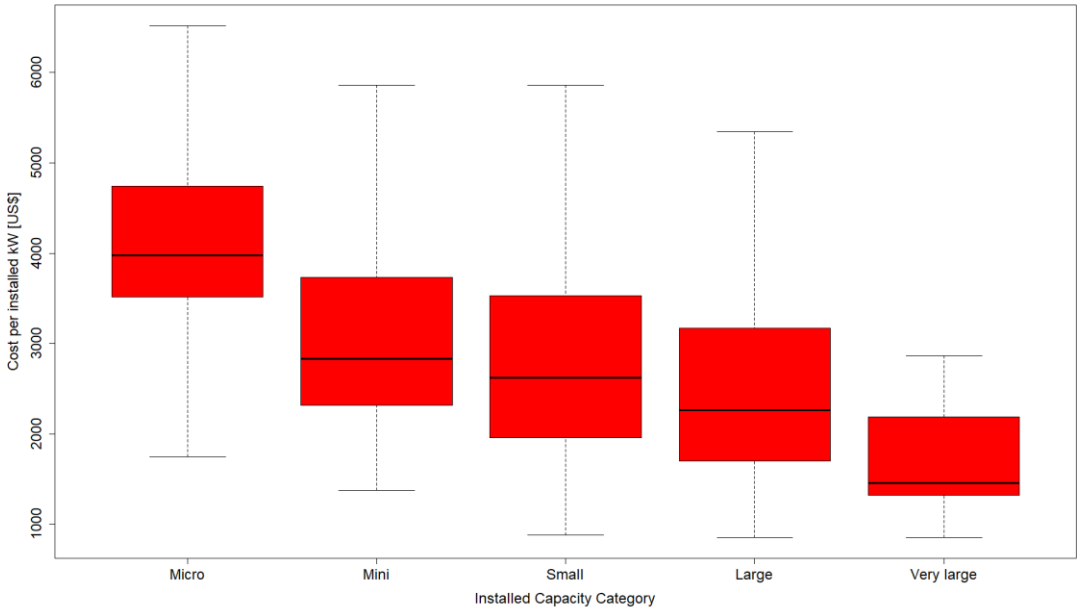


Figure 7.36: Costs per installed kW of all projects at IRR ≥ 5% by class in US$

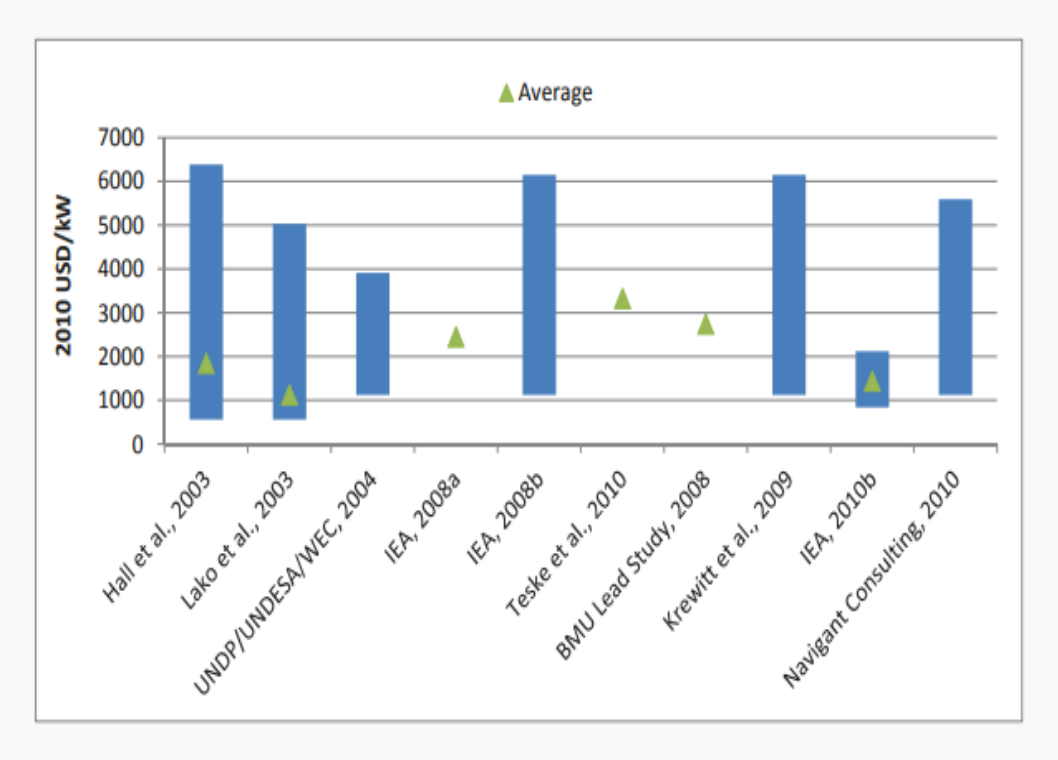


Figure 7.37: Costs (US$) per installed kW from various studies (IRENA, 2012)

Levelised Cost of Electricity (LCOE) is a common method of comparing differing methods of electricity generation. It is an economic measure of the total cost (capital cost and operating costs) and the energy generated over the project lifetime, considering the discount rate. The LCOE is also a measure of the minimum cost at which the electricity must be sold to break even and defined by:

$$\mathbf{LCOE} = \frac{\sum_{t=1}^{n} \frac{\mathbf{I_t + M_t + F_t}}{(\mathbf{1+i})^t}}{\sum_{t=1}^{n} \frac{\mathbf{E_t}}{(\mathbf{1+i})^t}} \qquad (7.3)$$

where t is the year of operation, n is the life time of the project [years], I_t are the capital costs in the year t (it is assumed all capital costs are incurred by year 0), M_t are maintenance and operation costs in the year t, F_t are fuel costs in the year t (assumed zero for hydropower), E_t is the energy generated in year t (kWh) and i is the discount rate (%) (e.g. 5% = 0.05).

Figure 7.38 shows the range of LCOE at three discount rates (5%, 10% and 15%) for all the projects, and at 10% discount rate for those projects with an IRR of over 10% and at 15% discount rate for those projects with an IRR of over 15%. LCOE within this model compares well to published LCOE costs although slightly high for China specifically (see figure 7.39). However, the costs published in the IRENA study are for projects with a lifetime of between 40 and 80 years as opposed to the 25 years assumed in this project. Therefore, the LCOE formula was applied again with 40 and 80 years lifetime (see figure 7.40) at 5% discount rate, which are closer to the published LCOE costs for China in the IRENA study. Furthermore, the costs published in the IRENA figures will include large hydro schemes with a reservoir and not just ROR schemes

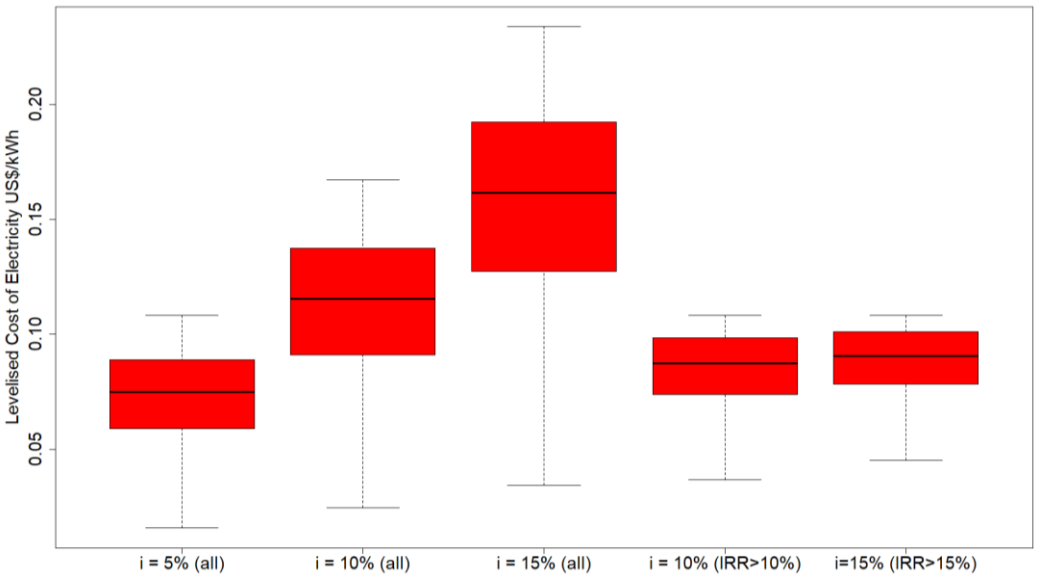


Figure 7.38: Levelised Cost of Electricity (LCOE) per kWh (US$) over the 25-year lifetime of all viable projects at 5%, 10% and 15% discount rate, plus at 10% discount rate for projects of IRR≥10% and 15% discount rate for projects at IRR≥15%

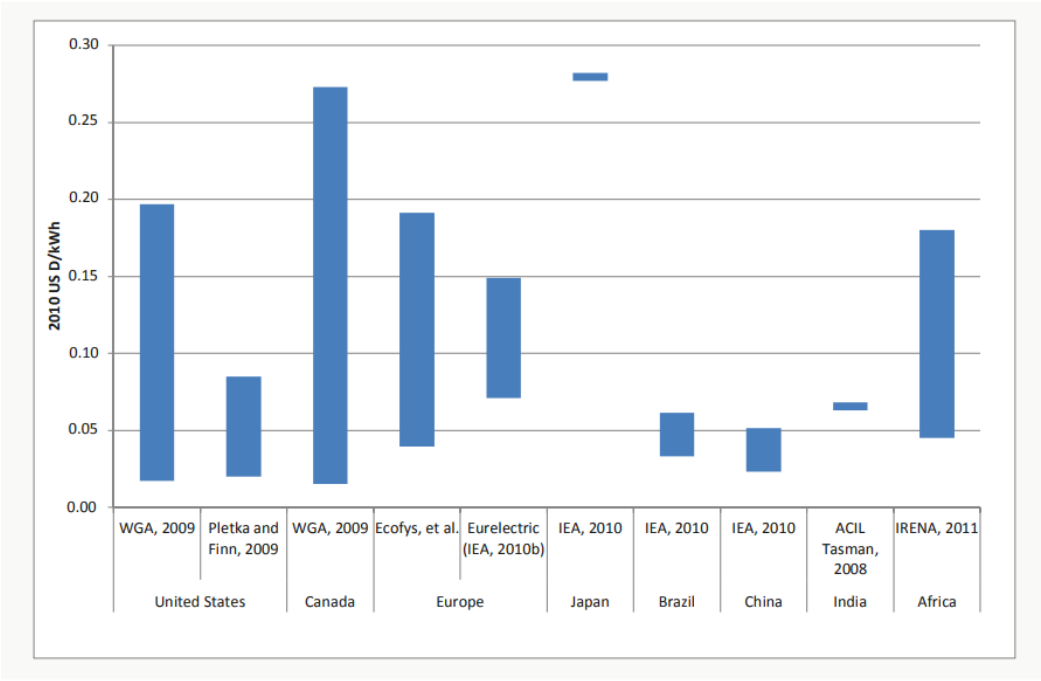


Figure 7.39: Levelised costs of electricity (LCOE) from various studies for hydropower installations by country (IRENA, 2012)

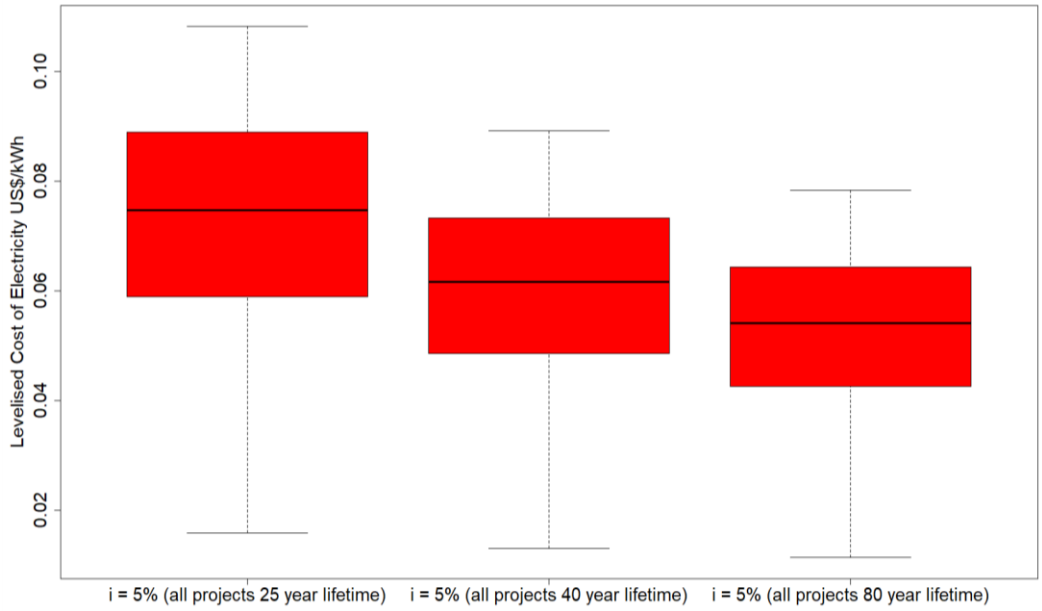


Figure 7.40: Levelised Cost of Electricity (LCOE) per kWh (US$) over the 25-year, 40-year and 80-year lifetime of all viable projects at 5% discount rate for all viable projects

A percentage breakdown of the costs by component (see figures 7.41 to 7.56) show that energy equipment, civil works, penstock and miscellaneous costs represent almost 80% of all costs at each installed kW category. For all projects, energy equipment represents a significant cost.

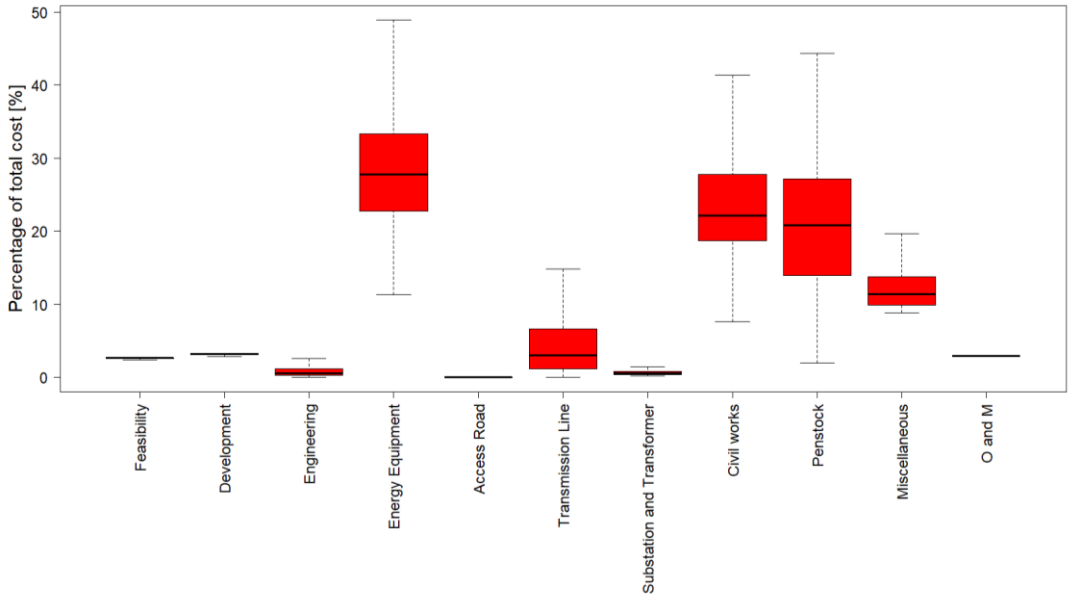


Figure 7.41: Percentage of costs by component for all projects with IRR ≥ 5%

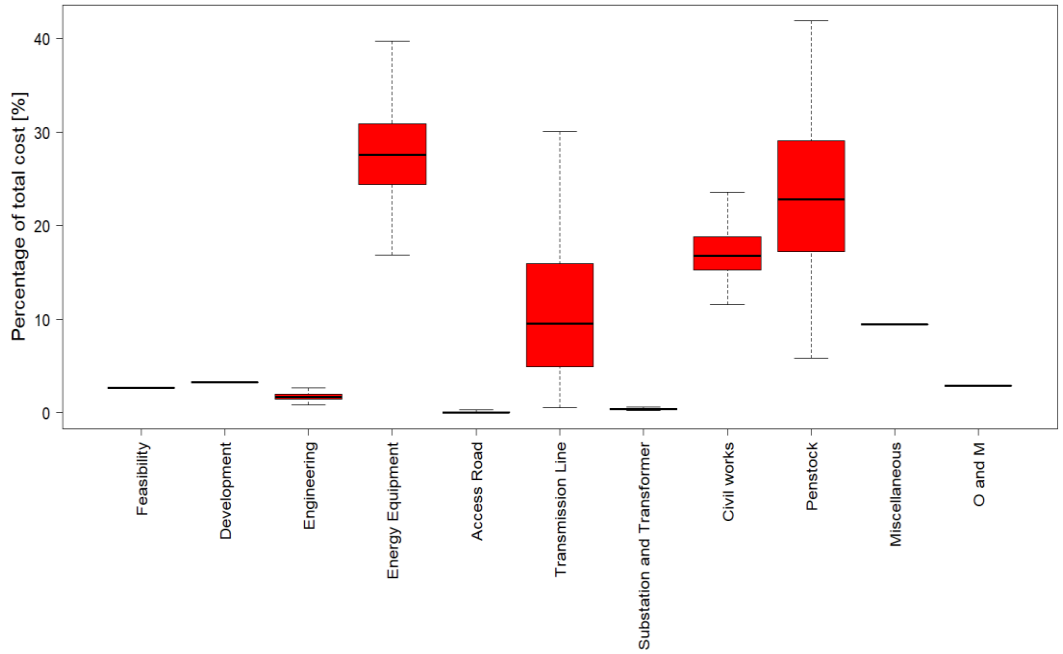


Figure 7.42: Percentage of costs by component for micro hydro projects with IRR ≥ 5%

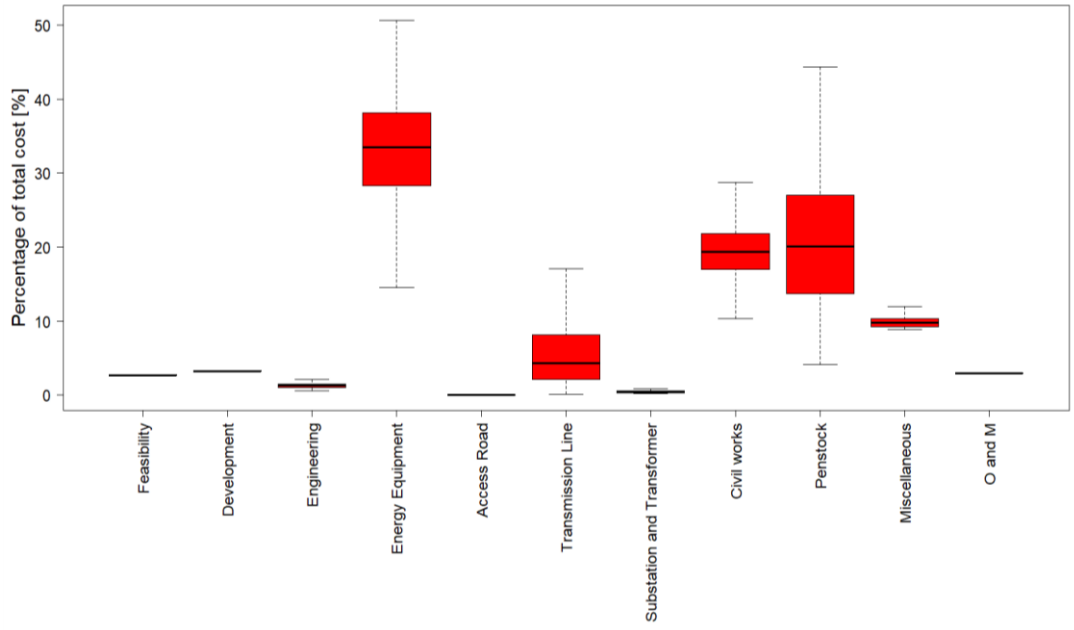


Figure 7.43: Percentage of costs by component for mini hydro projects with IRR ≥ 5%

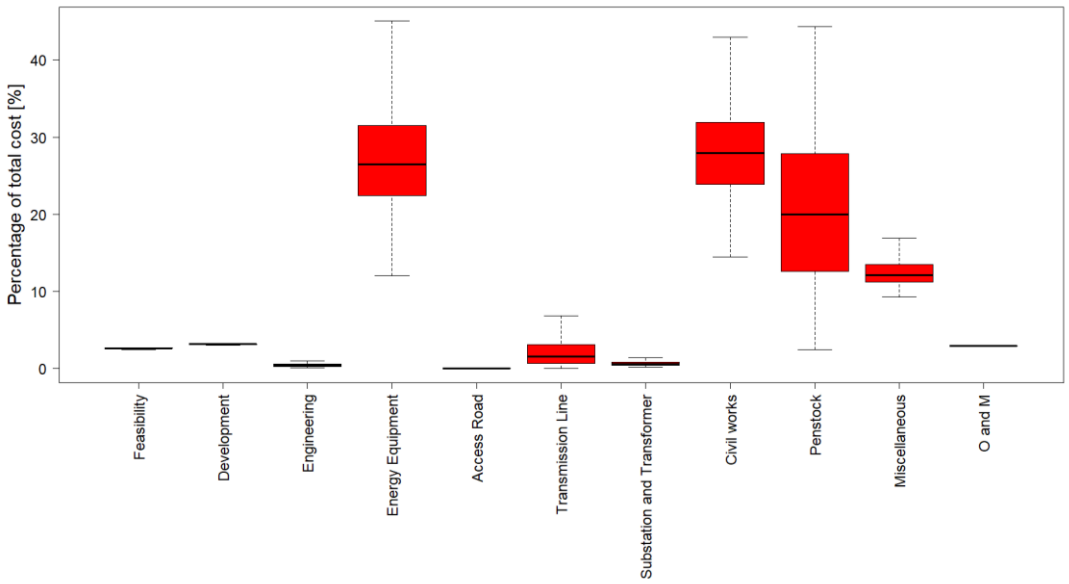


Figure 7.44: Percentage of costs by component for small hydro projects with IRR ≥ 5%

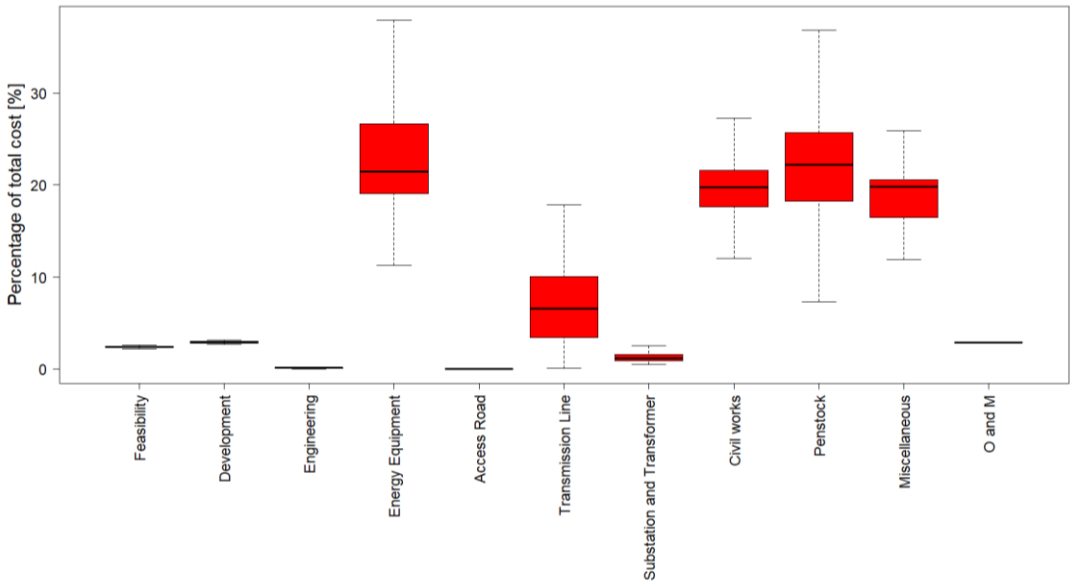


Figure 7.45: Percentage of costs by component for large hydro projects with IRR ≥ 5%

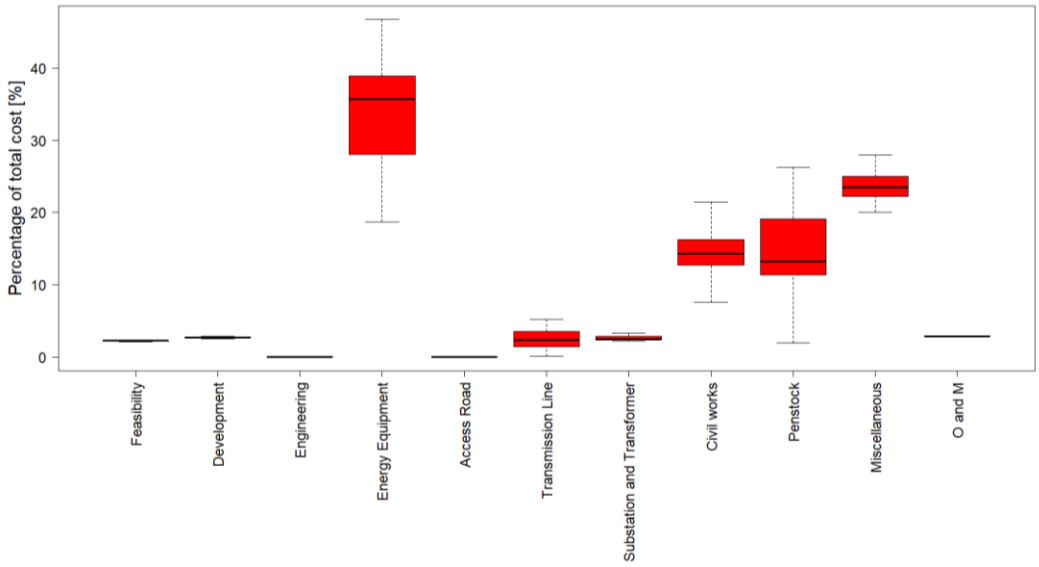


Figure 7.46: Percentage of costs by component for very large hydro projects with IRR ≥ 5%

7.5.1 Comparison of component costs to other studies

Studies of component costs are limited within the literature, and hence it is difficult to gauge the accuracy of the component cost contributions to the overall project costs. The IRENA Renewable Energy Technologies: Cost Analysis Series report (IRENA, 2012) reviews the component costs of 17 small hydropower stations ranging from 0.1MW to 24MW, expressed as a percentage by component and given in figure 7.47. It is important to note that the components themselves are expressed differently and the IRENA cost does not account for penstock, most likely as the penstocks are incorporated in the civil works costs and/or the equipment costings. In figure 7.44, the largest share of the components costs of the small hydropower plants are the energy equipment, civil works and penstock costs (median values of approximately 26%, 29% and 20% respectively – total 75%). In the IRENA study the greatest share of the component costs are equipment and civil works costs (median values of 25% and 33% respectively – total 58%).

At a glance it may then appear there is some over-estimation of equipment costing in the RETScreen analysis and under-estimation of civil works costs. The percentage share of civil works lowers as projects increase in size, which is contrary to what would be expected,

suggesting that the RETScreen method is not fully suitable for projects beyond the stated size of 50MW.

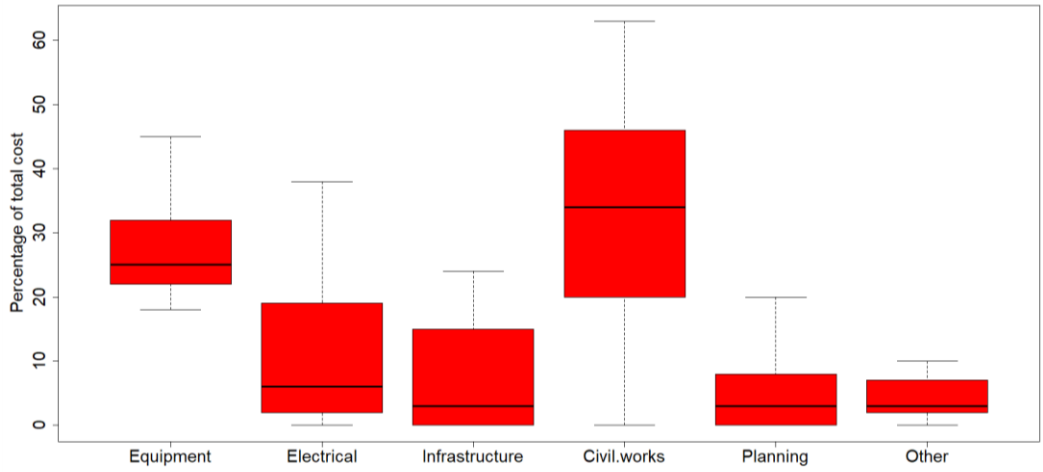


Figure 7.47: Percentage of costs by component small hydropower projects from the IRENA Renewable Energy Technologies: Cost Analysis Series (IRENA, 2012)

7.6 Cost sensitivity study

A cost sensitivity study of variables was carried out by re-analysing the costs of each viable project after increasing or decreasing one variable at a time by 10%, 25% or 50%. The new costs were then compared to the original costs, and the percentage change in costs per percentage change in variable was found by:

$$\Delta \mathbf{Y}\,\% = \frac{\textbf{New cost per kW or MWh after changing variable by} \pm \mathbf{X}\,\%}{\textbf{Oringal costs per kW or MWh}} \times \frac{\mathbf{1}}{\mathbf{X}} \qquad (7.4)$$

where Δ Y % is the percentage change in costs per percentage change in variable and X % is the percentage change in variable (10%, 25% or 50% increase / decrease).

These are plotted on four boxplots in figures 7.48, 7.49, 7.50 and 7.51, each with the same y-axis limits for easy comparison of variables. Figure 7.48 shows the percentage change in cost per kW after increase of all the variables by 5%, 10% and 50%. Labels on the x-axis represent [variable][percent][p][costkW] – where p represents when X% is a positive. Similarly, 7.49 shows the percentage change in cost per kW after decrease of all the variables by 10%, 25% and 50%, with x-axis labels as [variable][percent][n][costkW], where n represents negative

i.e. decrease. Figures 7.50 and 7.51 are the same as figures 7.48 and 7.49 except for change in cost per MWh.

From figures 7.48 and 7.49, the variables can be placed into high sensitivity ($\Delta Y\% > 0.5\%$ or $< -0.5\%$), medium sensitivity (between 0.2% and 0.5% or between -0.2% and -0.5%) or low sensitivity groups (between -0.2% and 0.2%) in terms of their impact on cost per kW (defined if the lower/upper quartile of any of the variables is within the group):

High sensitivity variables: Ccon (i.e. the inflation and exchange rate), flow, gross head, turbine efficiency.

Medium sensitivity variables: Penstock length, penstock diameter, design flow, Head-loss, F factor, generator and control costing equation, turbine and governor costing equation and the civil works costing equation.

Low sensitivity variables: B factor, river width estimation, access road length, transmission line length, energy available, income per MWh, installation of energy equipment costing equation, access road costing equation, transmission line costing equation, substation and transformer costing equation, penstock costing equation, installation of penstock costing equation, development costing equation, feasibility study costing equation, miscellaneous costing equation and the O and M costing equation.

Similarly, the sensitivity on impact on cost per MWh was also placed into groups:

High sensitivity variables: Ccon (i.e. the inflation and exchange rate), gross head, turbine efficiency, energy available, flow, design flow, penstock diameter.

Medium sensitivity variables: Penstock length, head-loss estimation, F factor, generator and control costing equation, turbine and governor costing equation and the civil works costing equation.

Low sensitivity variables: B factor, river width estimation, access road length, transmission line length, energy available, income per MWh, installation of energy equipment costing equation, access road costing equation, transmission line costing equation, substation and transformer costing equation, penstock costing equation, installation of penstock costing equation, development costing equation, feasibility study costing equation, miscellaneous costing equation and the O and M costing equation.

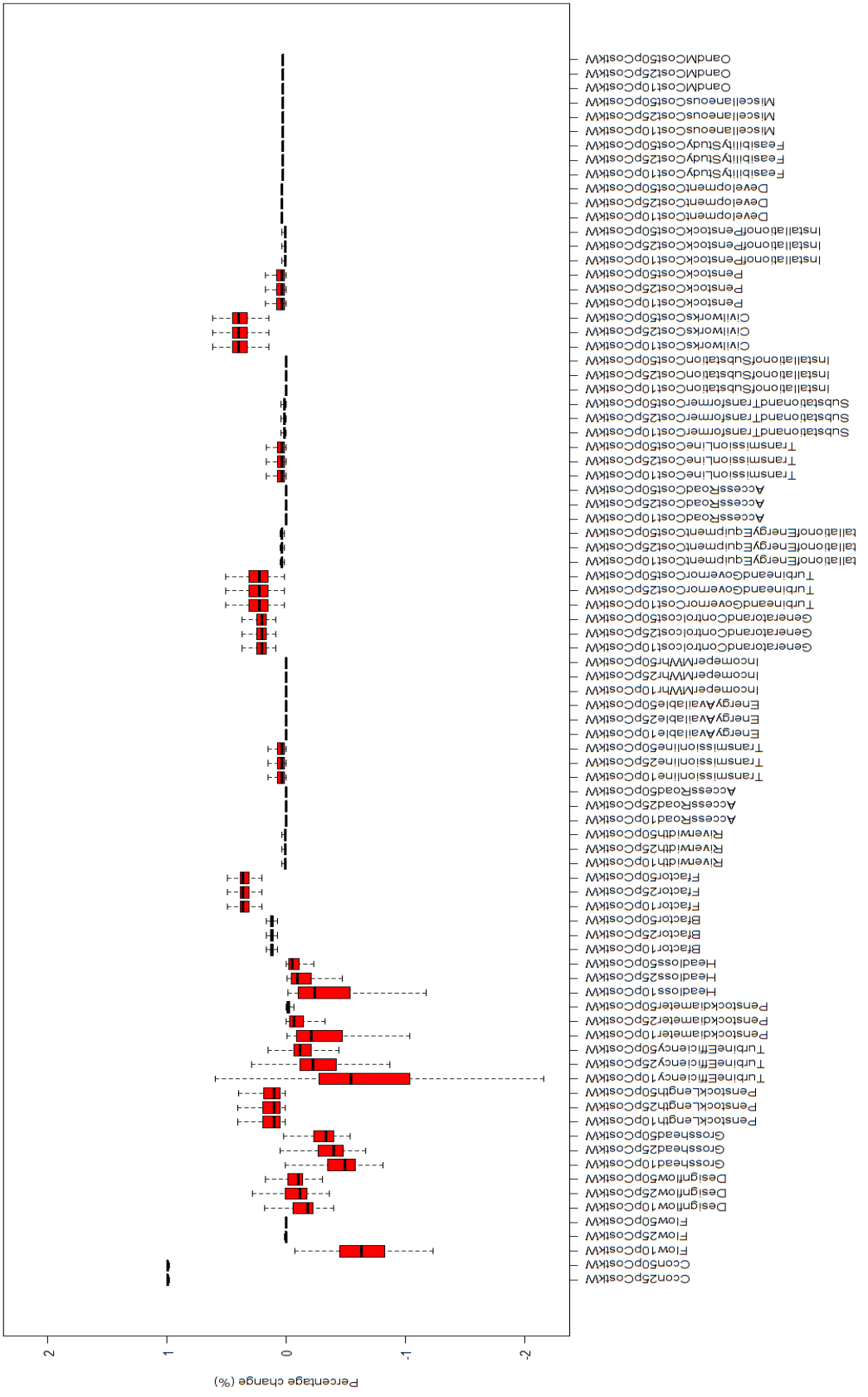


Figure 7.48: % change in cost per installed kW per % change in variable increase, with the x-axis labels in the form [variable][p][% change in variable][CostkW] where p = positive increase

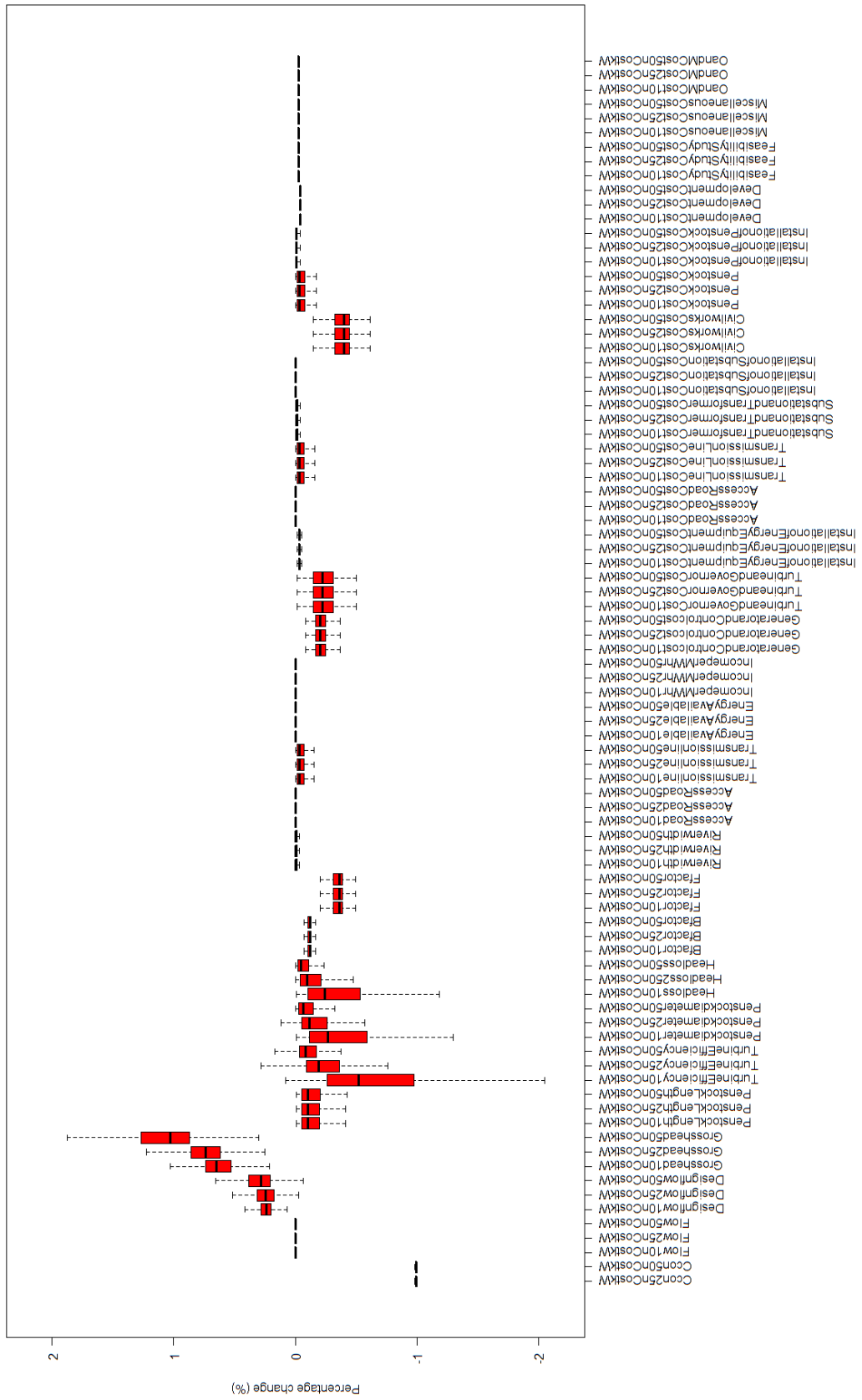


Figure 7.49: % change in cost per installed kW per % decrease in variable, with the x-axis labels in the form [variable][n][% change in variable][CostkW] where n = negative, decrease

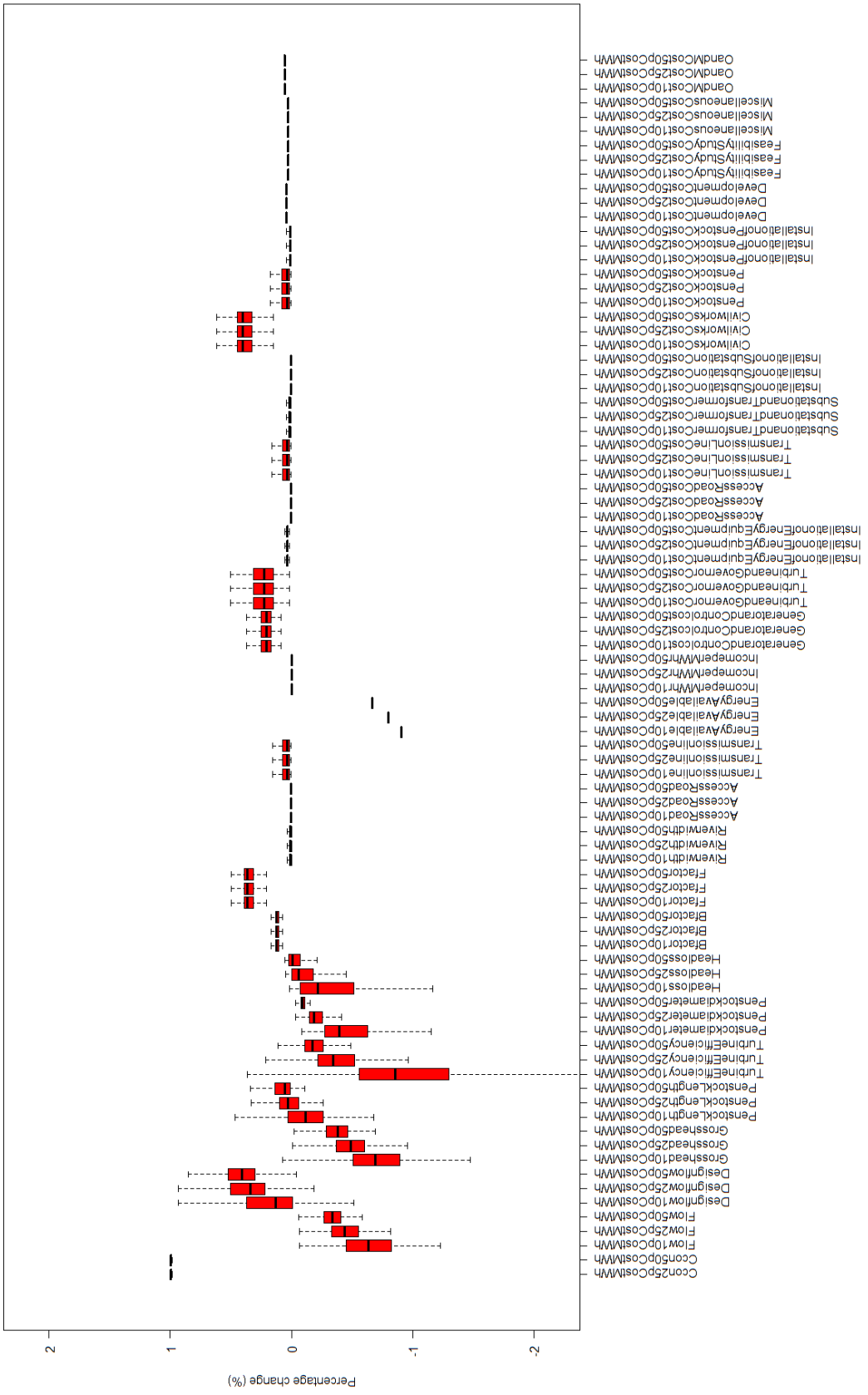


Figure 7.50: % change in cost per MWh per % increase in variable, with the x-axis labels in the form [variable][p][% change in variable][CostMWh] where p = positive, increase

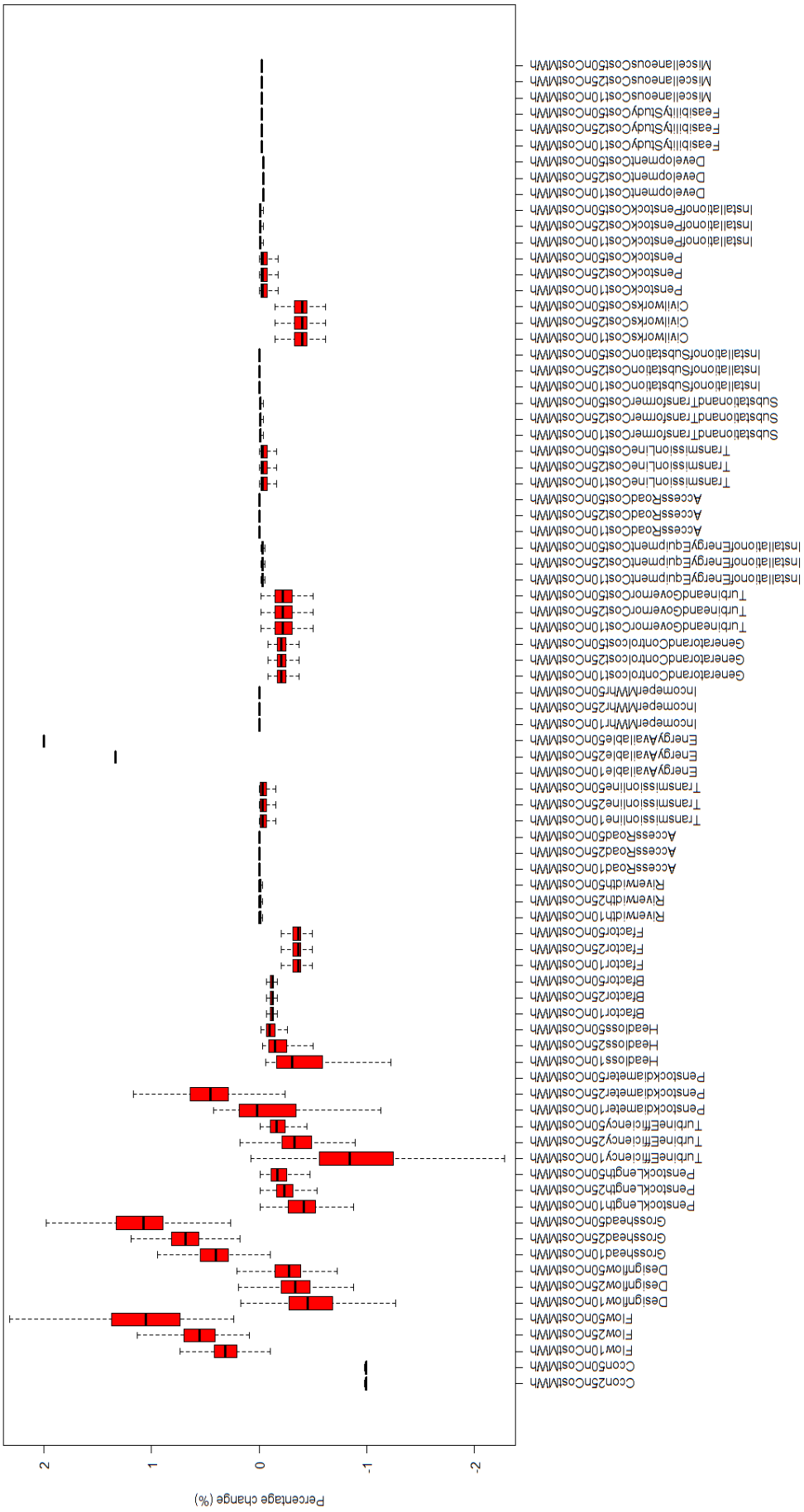


Figure 7.51: % change in cost per MWh per % decrease in variable, with the x-axis labels in the form [variable][n][% change in variable][CostMWh] where n = negative, decrease

7.7 Available power time-series

By rearranging the hydropower search algorithm, the daily power generated from each viable configuration can be established. Figure 7.52 represents the sum of all power generated (MWh) per day across the catchment using meteorological data from 2004. Available power decreases from January to April, increasing again between April and July to a peak which is then sustained at near maximum levels until September/October, and then decreases back to January levels by the end of December. Figure 7.53 is an x-y scatter plot of available power (MWh) per day for every viable configuration with IRR ≥ 5% during the year 2004.

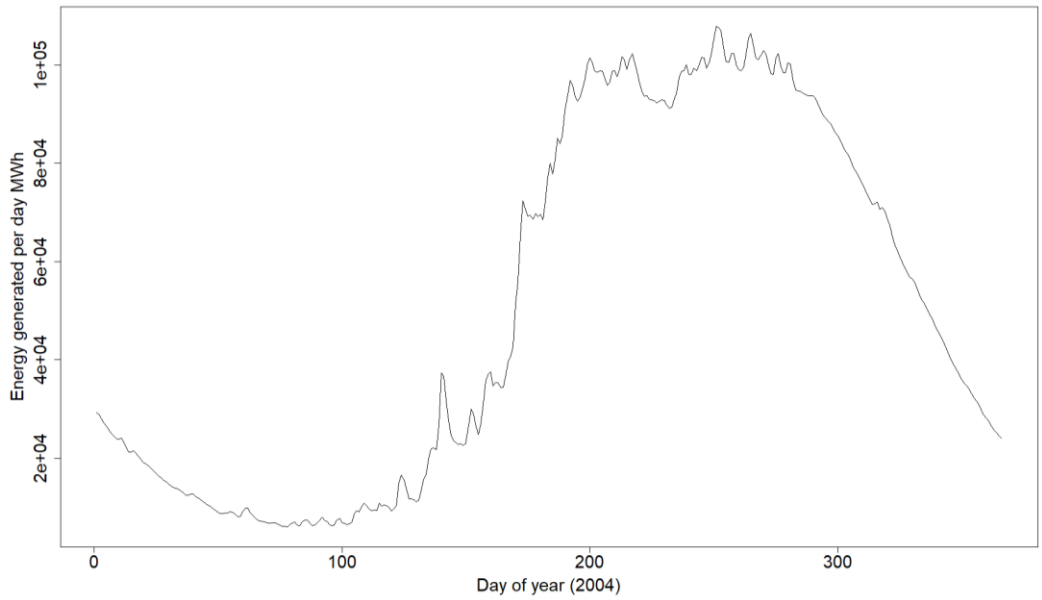


Figure 7.52: Sum of available power per day from all viable configurations with IRR ≥ 5% with flows from the year 2004

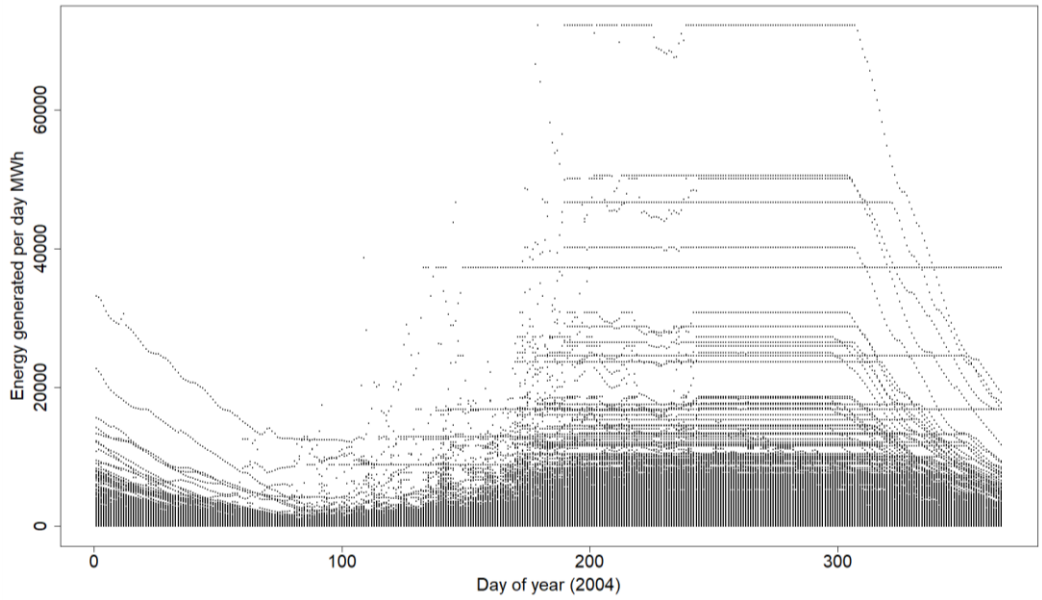


Figure 7.53: x-y scatter plot of available power per day from all viable configurations with IRR ≥ 5% with flows from 2004

Individual schemes are considered in figures 7.54, 7.55, 7.56, 7.57, 7.58 and 7.59 which represent the daily power available for selected example stations across the range of installed power capacity. The stations with higher installed plant capacity (figures 7.54, 7.55, 7.56 and 7.57) exhibit a similar pattern of flow to the power generated by the catchment as a whole, shown in figure 7.52. However, the smaller stations (figures 7.58 and 7.59) exhibit intermittent power generation across the year, either generating maximum power or power well below maximum (or even none). This is particularly noticeable in figure 7.59 where the turbine is only generating power for brief periods following storm events.

It must be remembered that the algorithm was designed to search for configurations that would generate the maximum NPV, and not necessarily the most energy or the most stable supply. However, an operator may prefer a scheme operating at a lower Q selection (and hence lower design flow) to provide a more stable supply to the grid with the station operating at a higher plant capacity factor, possibly generating less overall power or profit. The algorithm would need only a small adjustment to provide this flexibility.

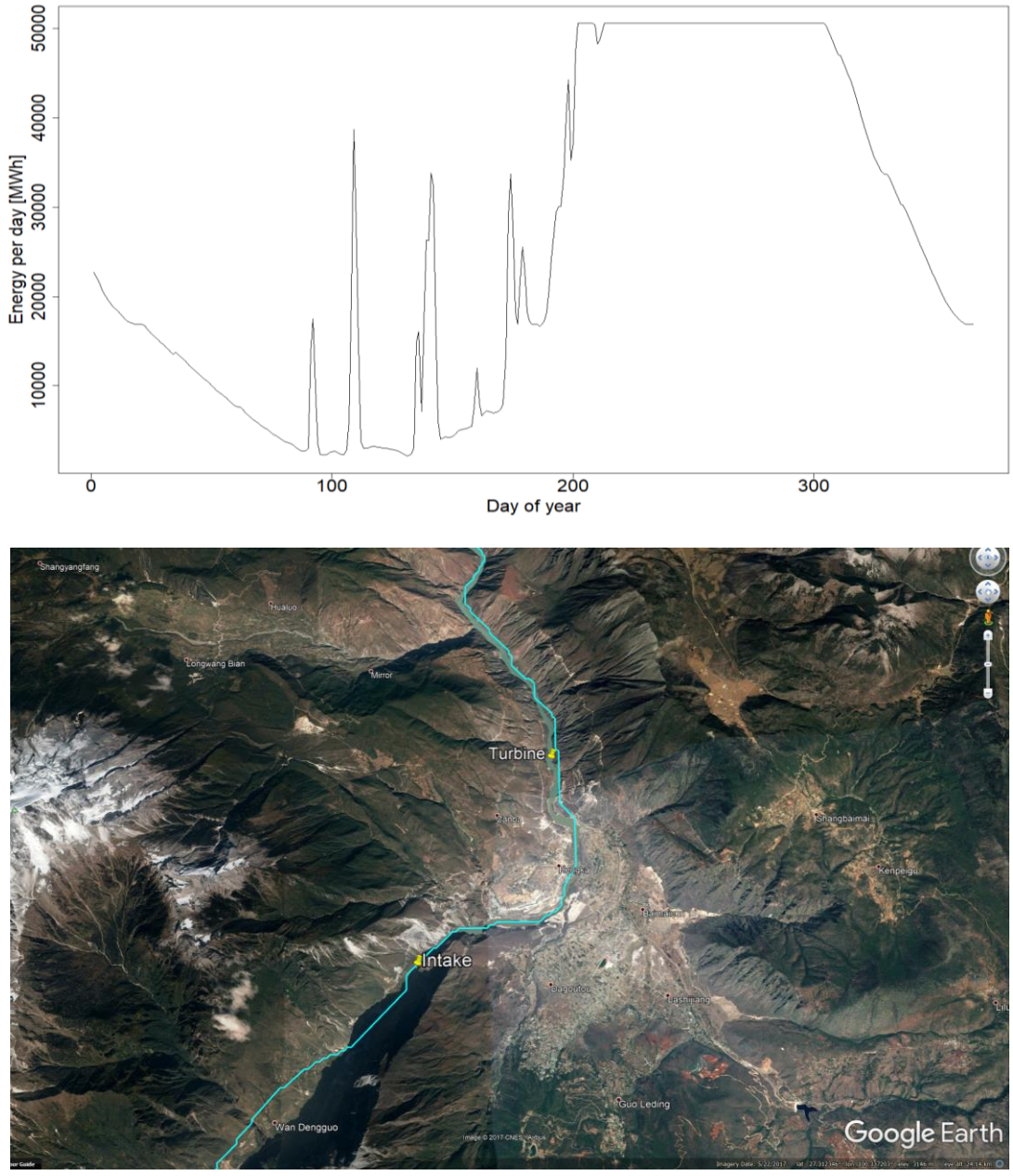


Figure 7.54: Station 2 energy generated (MWh) per day in 2004 (top) and a Google Earth representation showing the intake and turbine locations (bottom)

Scheme details: 2.2 GW installed capacity, design flow 1,827 m^3s^{-1}, gross head =146 m, QSelection=Q_{25}, river width = 425m, intake elevation = 1728m, turbine = 3* Francis, energy available per year = 7.84TWh, plant capacity factor = 0.4, penstock length = 8.6km, transmission line length = 53km, transmission Line kV = 1100, IRR = 29.2%

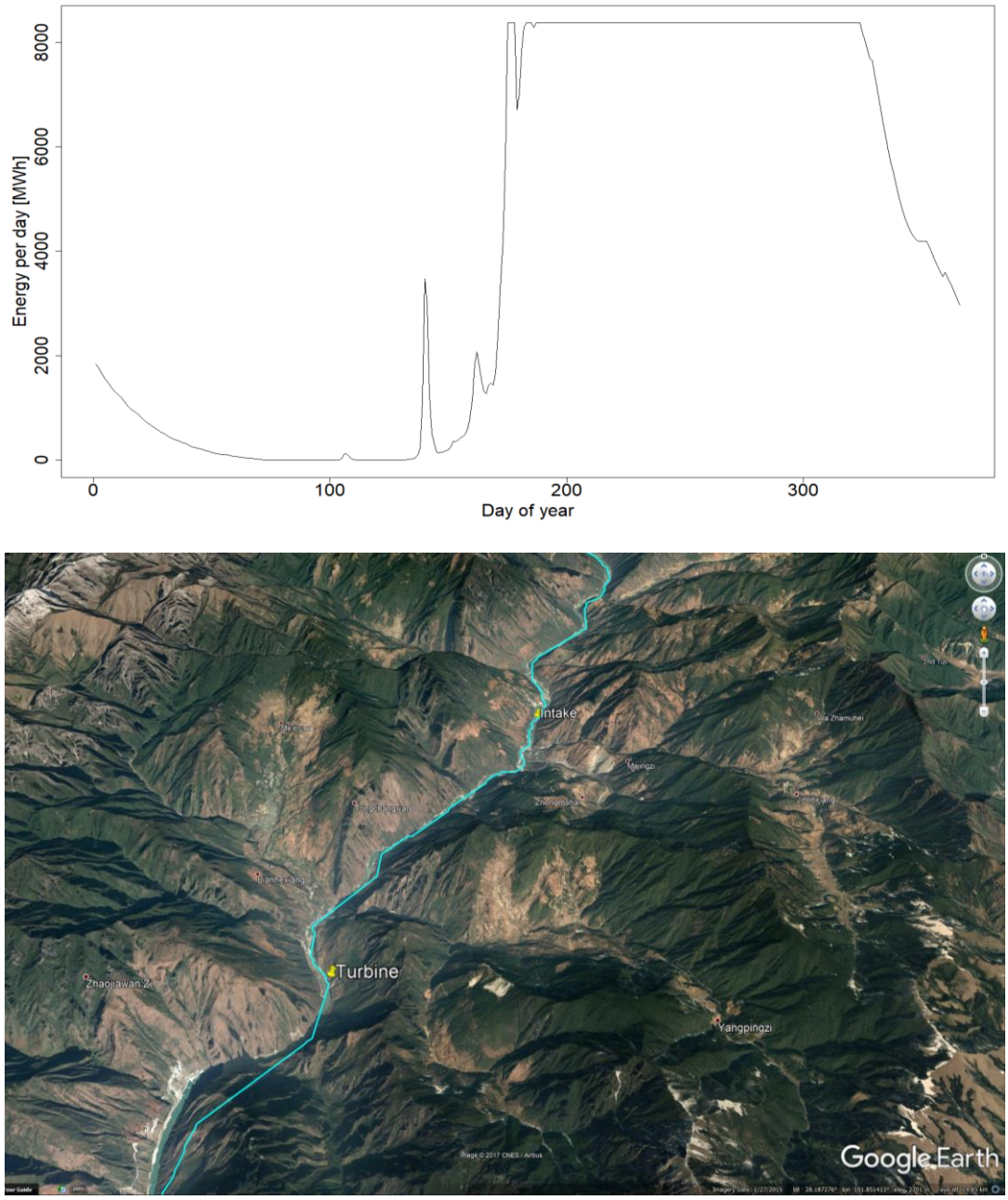
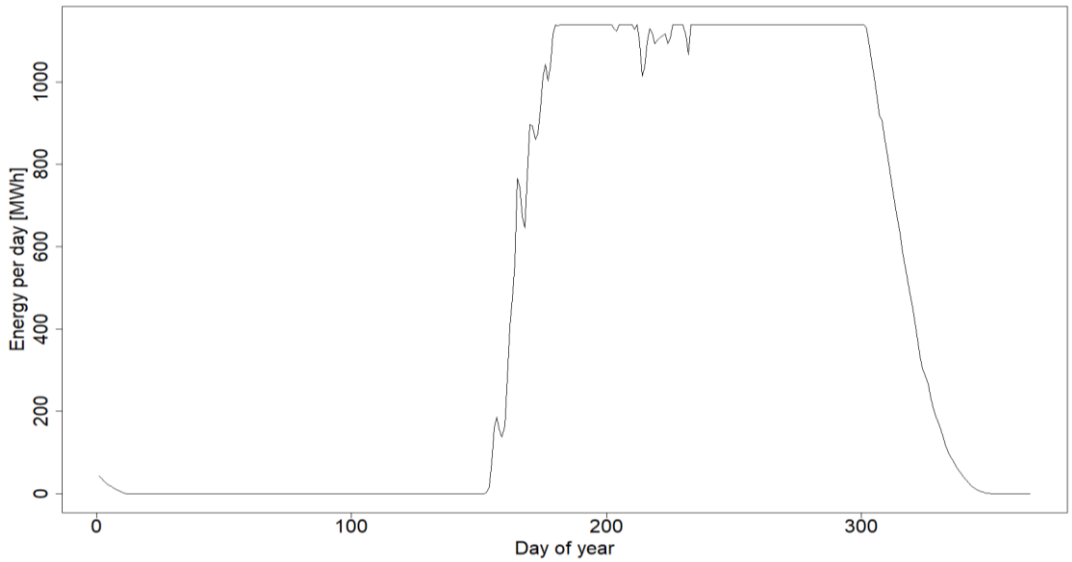


Figure 7.55: Station 100 energy generated (MWh) per day in 2004 (top) and a Google Earth representation showing the intake and turbine locations (bottom)

Scheme details: installed capacity=369MW, design flow=1,438m^3s^{-1}, gross head =32m, QSelection=Q_{35}, river width = 445m, intake elevation=1353m, turbine = 2* Francis, energy available per year=1.37TWh, plant capacity factor = 0.42, penstock length=11.2km, transmission line length = 53km, transmission Line kV=1100, IRR=15.8%



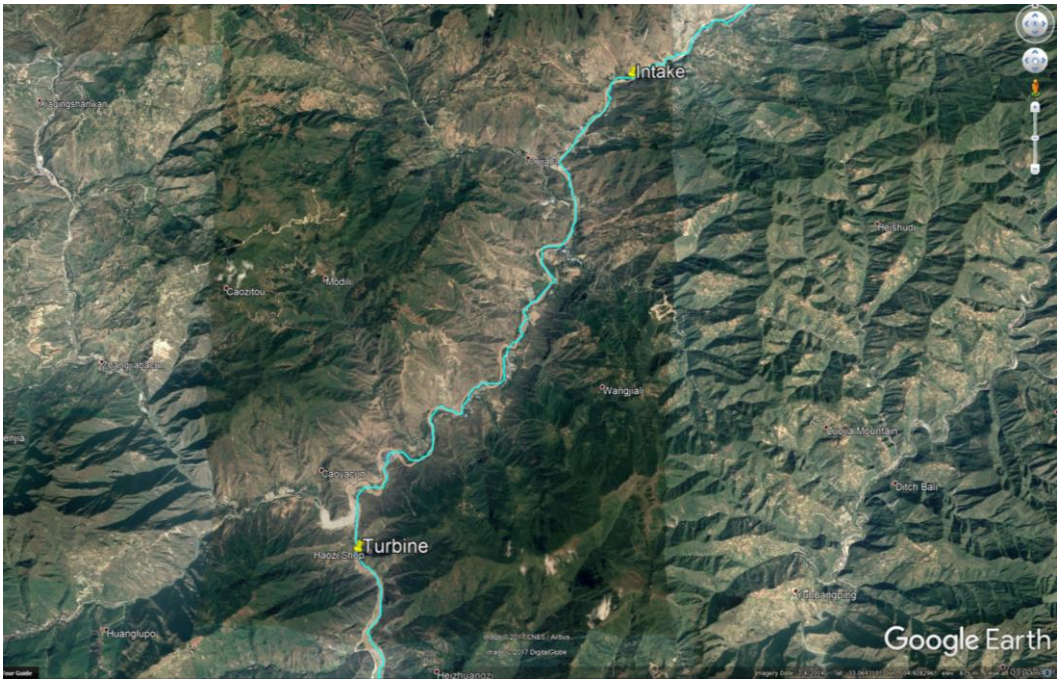
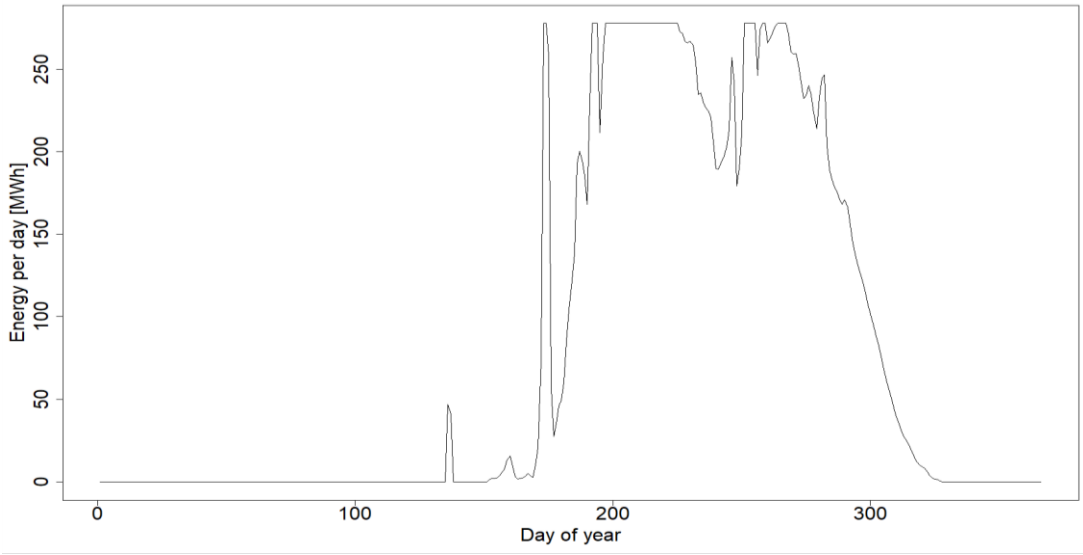


Figure 7.56: Station 390 energy generated (MWh) per day in 2004 (top) and a Google Earth representation showing the intake and turbine locations (bottom)

Scheme details: installed capacity=46MW, design flow=109m^3s^{-1}, gross head=52m, QSelection=Q_{25}, river width=120m, intake elevation=832m, turbine=1* Francis, energy available per year=163GWh, plant capacity factor = 0.41, penstock length=14.9km, transmission line length=3.6km, transmission Line kV=200, IRR=16.5%



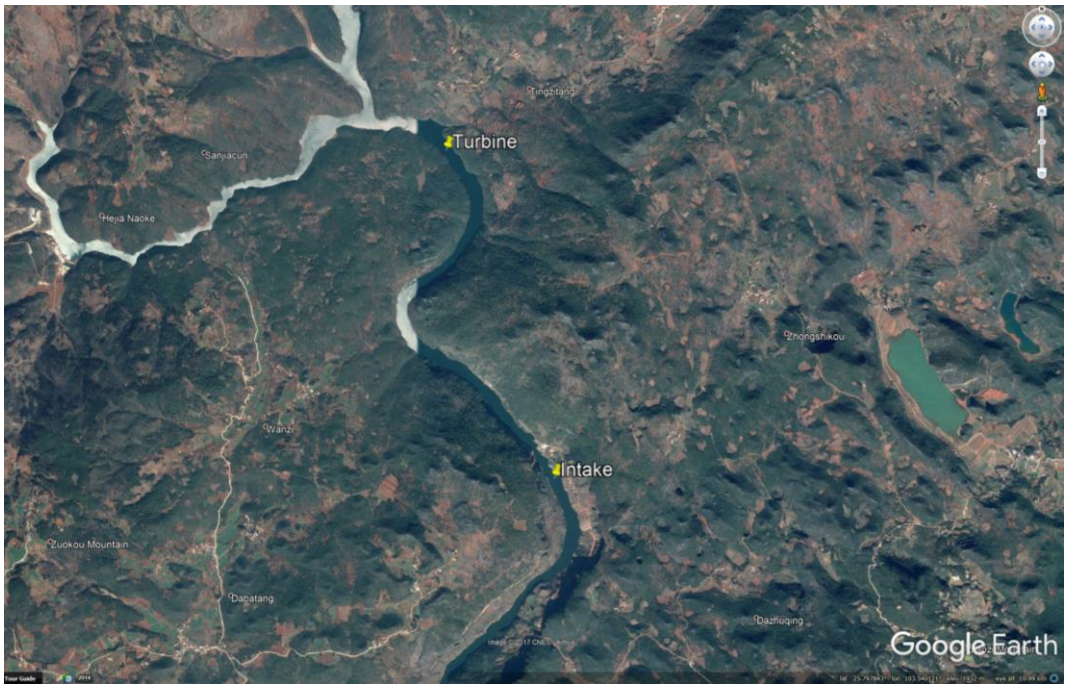


Figure 7.57: Station 970 energy generated (MWh) per day in 2004 (top) and a Google Earth representation showing the intake and turbine locations (bottom)

Scheme details: installed capacity=11MW, design flow=41m^3s^{-1}, gross head=34m, QSelection=Q_{15}, river width=60m, intake elevation=1178m, turbine=1* Francis, energy available per year=29GWh, plant capacity factor = 0.29, penstock length=3.6km, transmission line length=14.5km, transmission Line kV=66, IRR=9.7%

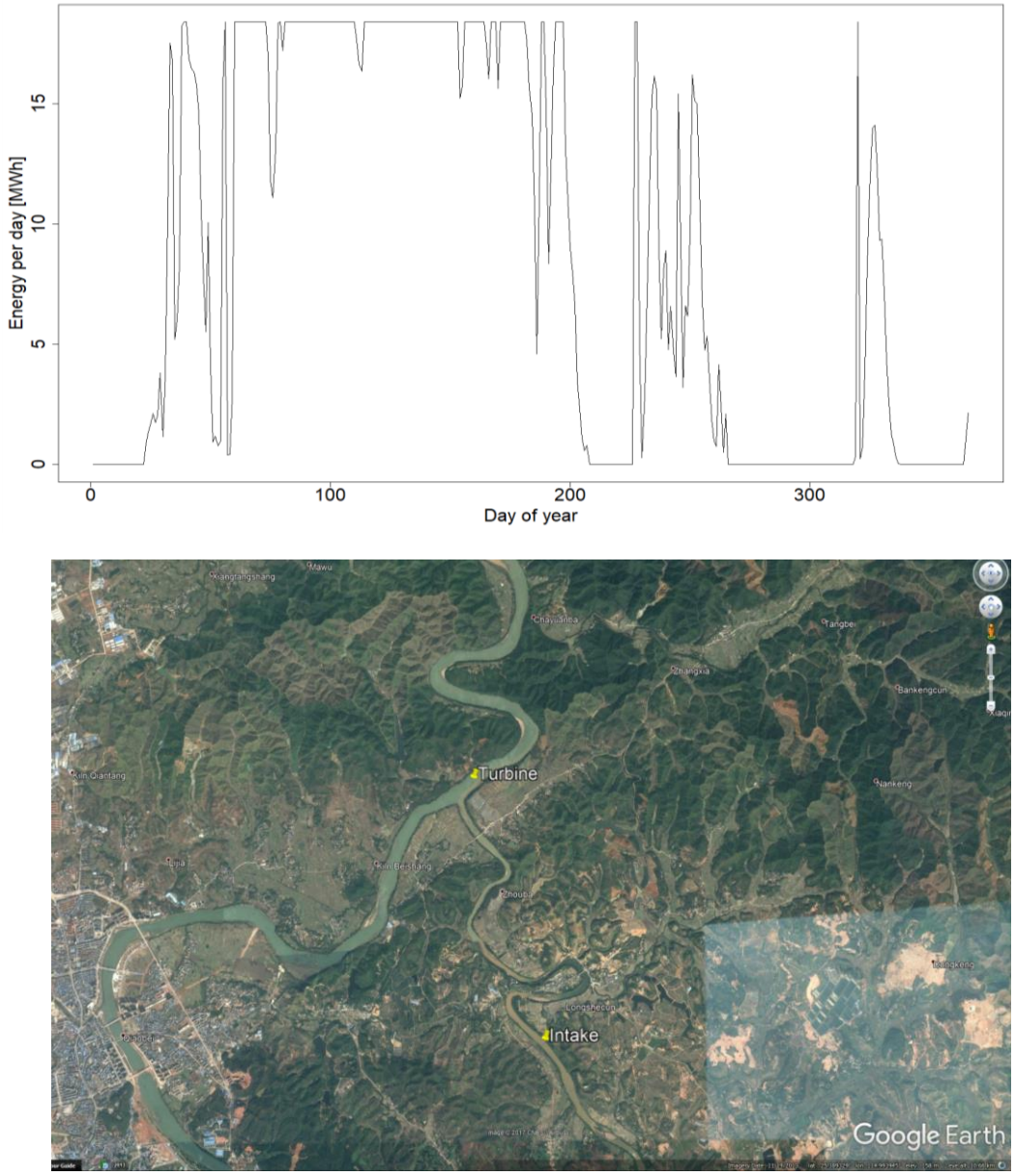


Figure 7.58: Station 1591 energy generated (MWh) per day in 2004 (top) and a Google Earth representation showing the intake and turbine locations (bottom)

Scheme details: installed capacity=1.5MW, design flow=8.4m^3s^{-1}, gross head=23m, QSelection=Q_{30}, river width=62m, intake elevation=143m, turbine=1* Francis, energy available per year=5.3GWh, plant capacity factor = 0.42, penstock length=3.6km, transmission line length=6.6km, transmission Line kV=11, IRR=8.2%

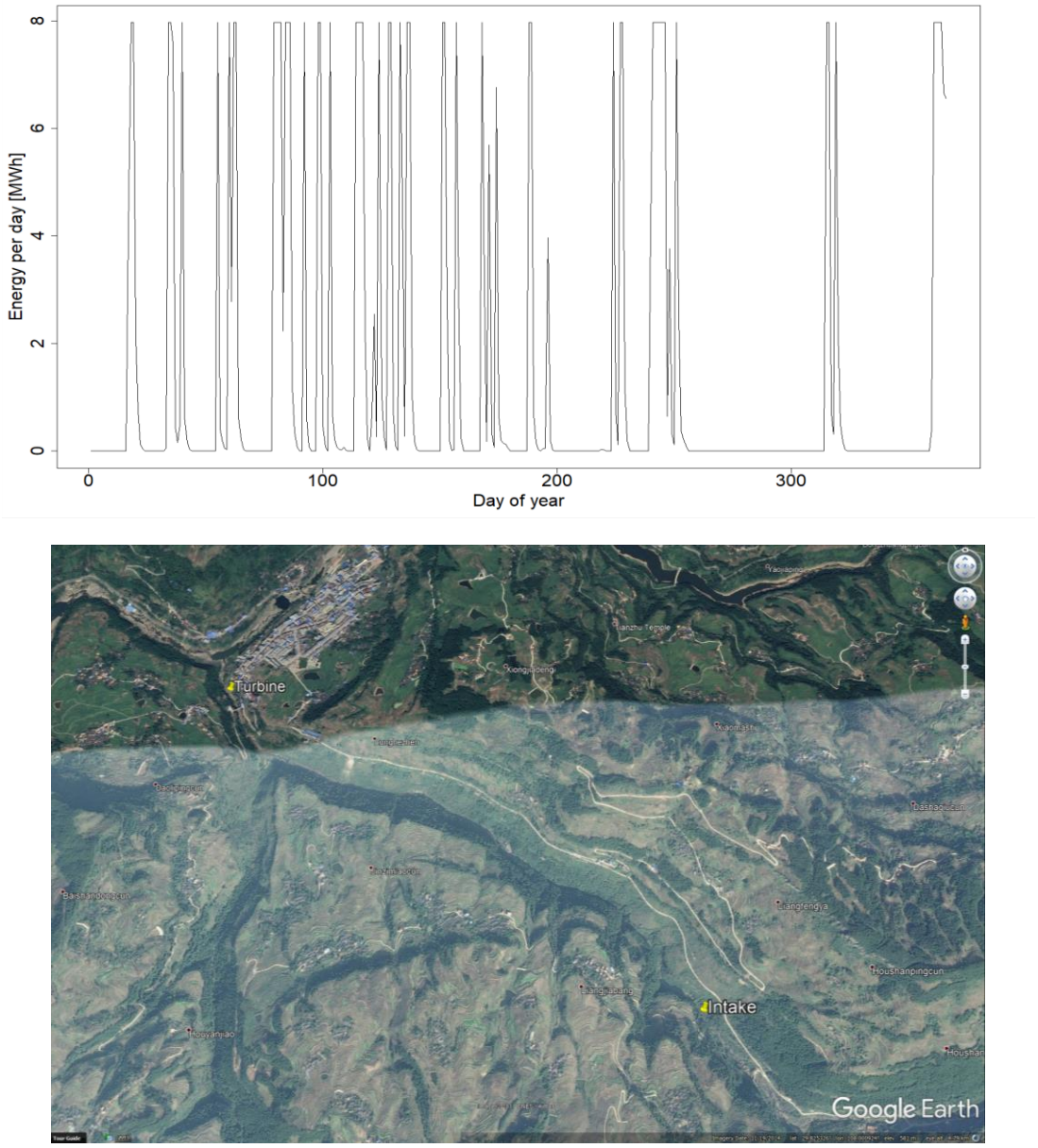


Figure 7.59: Station 2400 energy generated (MWh) per day in 2004 (top) and a Google Earth representation showing the intake and turbine locations (bottom)

Scheme details: installed capacity=132kW, design flow=0.09m^3s^{-1}, gross head=175m, QSelection=Q_{20}, river width=13.1m, intake elevation=561m, turbine=1*Pelton, energy available per year=462MWh, plant capacity factor = 0.40, penstock length=3.6km, transmission line length=9.8km, transmission Line kV=11, IRR=8.2%

7.8 Plant capacity factors

Plant capacity factors (the amount of power generated compared to the theoretical maximum output) range from 0.1 to 0.7 (see figure 7.60), with a majority between 0.2 and 0.5. This is due to the configurations being designed to take advantage of storm flows to yield high power although only over small periods of time, and hence maximise NPV.

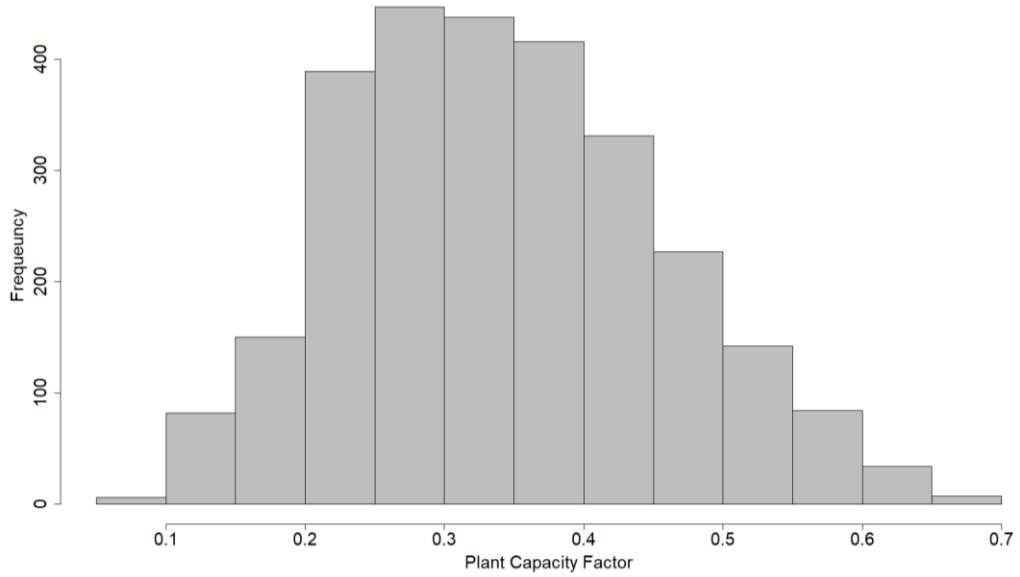


Figure 7.60: Histogram of Plant Capacity Factors

7.9 Further investigation of access road length

As mentioned earlier, it was a surprise to see the very high frequency of access roads with a very small length (95% of the access roads have a length less than 100m). Due to the high density of the original access road shapefile, the dataset was reduced to points at the centre of each grid square. Therefore, it is possible that the nearest access road to a schemes intake/turbine could be up to half a grid square diagonal further than the distance to the point in the centre of the grid. Therefore, the access road cost of each viable configuration was re-assessed adding 300m (i.e. half a grid square) to the original calculated length, and the impact this had on the overall IRR was calculated. The impact on IRR is minimal, meaning the simplification of the access road dataset to a point file was acceptable. However, 32 of the sites were affected to the point where the new IRR% dropped below 5%, 30 of which were

micro hydro sites and 2 were mini hydro sites, reducing the overall installed capacity by 2.8MW over the whole of the Yangtze.

7.10 Further investigation of transmission line voltage

Another potential problem with the original costings was the assumption that configurations with a transmission line voltage requirement of above 1100kV could be transmitted with a single 1100kV transmission line. The 141 configurations with a transmission line of 1100kV or greater were therefore re-costed as follows:

$$\textbf{New transmission line cost} = \frac{\textbf{Original cost} * \textbf{Required kV}}{\textbf{1100kV}} \qquad (7.5)$$

The impact on IRR of this re-costing is given in figure 7.61, and is quite large (with a median value of -12%). Of the 141 schemes with a transmission voltage requirement of 1100kV or more, the IRR of 28% of them would drop below 5% representing an installed capacity of 13GW. Hence, future assessments should consider the additional cost of high voltage transmission line requirements.

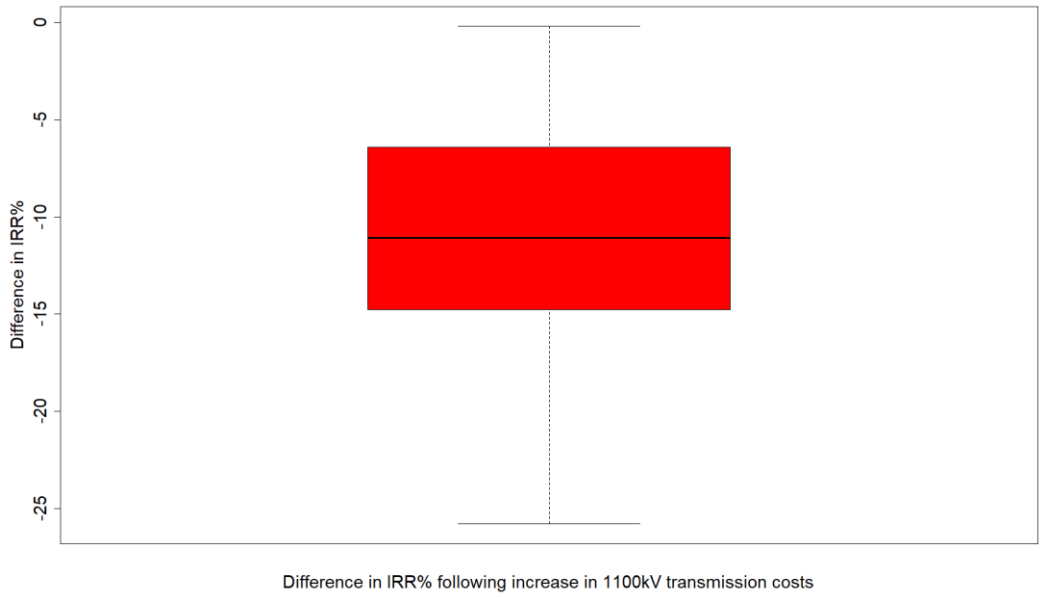


Figure 7.61: Boxplot of impact on IRR% by multiplying the transmission line cost by a factor of $\frac{Required\ voltage\ kV}{1100\ kV}$ *for sites with a required voltage above 1100kV*

7.11 Chapter conclusions

- 1,411,299 intake (river) locations across the YDB were investigated extending penstocks downstream up to a maximum of 15km, a total of approximately 276 million searches. Approximately 24 million (9%) were deemed viable with an IRR ≥ 5%. The configurations with the best NPV were selected leaving 2,100,609 configurations with an IRR ≥ 5%.
- The hydropower conflict algorithm was run at 5%, 10% and 15% discount rate, resulting in 13,604/8,127 and 4,534 viable run-of-river hydropower stations respectively with an installed capacity of 1.53TW/1.44TW and 1.21TW respectively.
- As discount rate is increased, the number of stations falls markedly, but the installed capacity less so, since many of the large power stations were still viable whatever the discount rate.
- The gross head as calculated by the model was compared to the gross head found visually within Google Earth, with a median overestimate of 40m which may have grossly inflated the amount of available hydropower.
- Due to time and computing constraints it was not possible to re-run the hydropower search algorithm at higher DEM resolution, but a novel method was developed to automatically underestimate the gross head from a higher resolution DEM. When this was compared to Google Earth, the median difference was negligible.
- Using the new calculated gross head, the hydropower conflict algorithm was run at 5%, 10% and 15% discount rate, resulting in 2,825/1,509 and 701 viable run-of-river hydropower stations respectively with an installed capacity of 130GW/103GW and 76GW respectively.
- The viable hydropower stations have a wide range of characteristics. A large proportion are single turbine schemes (dominated by Francis and Pelton type) with gross heads below 100m with design flows below $10 m^3 s^{-1}$ at between Q_{15} and Q_{30}.
- Most of the project costs estimated by the model were between 1000$ and 7000$ per installed kW, with median values between 2000$ and 4000$ per installed kW, decreasing as projects increase in size. This compares well with actual costs published by multiple studies within the literature.

- LCOE costs were estimated to be between 0.01 and 0.25US$ per kWh and with median results of between 0.075 and 0.15US$ per kWh. This is also in the range of costs published elsewhere within the literature.
- The highest proportion of the costs by component were shared between the energy equipment, penstock and civil works. There are similarities to the share of costs by component published elsewhere but also differences, suggesting the RETScreen model may be unsuitable for schemes beyond 50MW.
- In a cost sensitivity analysis, the inflation and exchange rate, flow duration curve, gross head, turbine efficiency and penstock diameter had the greatest influence on overall costs. This highlights the importance of the hydrological modelling, the use of a high quality and high-resolution DEM. Penstock diameter was not adjusted within the model to investigate the trade-off of head loss vs reduced cost of penstock and turbine efficiency was only calculated as per the RETScreen equations.
- A hydropower configuration as determined by the model would generate significantly more energy between April and July than other times of the year due to late spring/summer rains.
- Plant capacity factors ranged from 0.2 to 0.5, in part due to the prioritisation of NPV rather than a constant energy supply.
- The simplification of the road dataset may have shortened road lengths, reducing costs but this would have insignificant impact on IRR for the majority of the sites.
- Large schemes may require multiple 1100kV lines to be costed, which were not accounted for in this algorithm.

8 Conclusions

8.1 Introduction

This study aimed to answer the research question:

How can we develop a model to search for economical ROR hydropower resource across large river catchments, using the Yangtze drainage basin as a test case?

In achieving this, a gridded, distributed hydrological model was developed incorporating high-resolution historic meteorological datasets, with view to establishing the flow duration curves across the river network of the Yangtze River drainage basin as a test case catchment. The flow duration curves established were included in a new geospatial tool to test the energy generated and financial performance of potential ROR hydropower stations within the network. A conflict algorithm was developed to find the best performing hydropower network from across the catchment.

8.2 Summary of the work

Chapter 1 gives an introduction, setting the context and the importance of renewable energy supply development, sets the objectives of the work and the contribution to knowledge.

Chapter 2 presents background information relative to the aims of the project and a literature survey. The chapter presents information about primary energy consumption trends, climate change, characteristics of hydropower, hydropower assessments to date and costs of hydropower. The chapter also introduces the Yangtze River basin, hydrology, meteorology and GIS.

Chapter 3 defines the important input datasets required for a hydropower search algorithm and details the selection and/or development of a DEM, identification of the YDB, a lakes and reservoirs dataset, river width and depth estimation methodology, an access road dataset and an electrical transmission dataset. The chapter also identifies the need for development of a hydrological model to determine surface runoff across the catchment and a costing methodology.

Chapter 4 presents the development of historical meteorological datasets required as input to the hydrological model including precipitation, temperature and evapotranspiration datasets and a snowmelt model.

Chapter 5 presents the development of a hydrological model based on the G2G model incorporating the meteorological datasets developed in chapter 4. Calibration was carried out at 15-arcseconds resolution on sub-catchments of the YDB and compared to observed data by assessing the performance of 8 objective functions. The best performing parameter set across multiple sub-catchments was used to extract surface runoff data across the entire YDB.

Chapter 6 presents the development of a hydropower search algorithm incorporating the surface runoff time series output from the hydrological model and all the datasets from Chapter 3. To search for economically viable hydropower schemes, trial schemes were costed using the RETScreen methodology.

Chapter 7 presents the results output from the hydropower search, and concludes that a hybrid 15-arcsecond and 3-arcsecond DEM is most likely to produce accurate results. The characteristics of viable hydropower sites are analysed, and a cost sensitivity study conducted.

8.3 Main conclusions

- Primary energy consumption is rising dramatically, particularly in the developing world and closely linked with climate change. A reliable hydropower search tool across remote catchments would be attractive to many agencies concerned with energy development.
- This study witnessed the development of a ROR hydropower search algorithm, trialled on the 2 million km^2 Yangtze Drainage Basin, one of the largest catchment areas in the world that goes beyond any other tool published within the literature in terms of its scale and detail. This could prove to be a powerful tool for those parties interested in hydropower development.
- The hydropower search algorithm found an optimised catchment configuration of 2,825/1,509 and 701 viable and non-conflicting run-of-river hydropower stations at 5%/ 10% and 15% discount rate, with an installed capacity of 130GW/103GW and 76GW respectively.

- The hydropower configuration would generate significantly more energy between April and July than other times of the year.
- The algorithm conducted 276 million iterative searches, of which approximately 9% were deemed economically viable i.e. with an IRR of ≥ 5%.
- The range of installation costs per installed kW predicted by the algorithm compared well to published costs of existing hydropower stations.
- Estimated LCOE costs were also closely aligned with costs published elsewhere in the literature, adding strength to the economic performance of the algorithm.
- Costings and power production were based on the RETSCreen Small Hydro Project model. Energy equipment and civil works costs formed the highest share of component costs both in this model and studies published elsewhere in the literature. However, there appeared to be some under-estimation of civil works costs and over-estimation of energy equipment costs particularly as hydropower schemes became large. This could suggest that the RETScreen model deviates from reality, and not fully suitable, for schemes above the stated maximum of 50MW.
- In a cost sensitivity study, the costings were particularly sensitive to inflation rate, flow duration curve, gross head, turbine efficiency and penstock diameter. The penstock diameter and turbine efficiency calculations relied entirely upon the RETScreen calculations and hence were not manipulated. The gross head and hydrological model also developed for this project however relied upon the resolution and accuracy of the underlying DEM which was selected at 15-arcseconds as a compromise between model quality and runtimes.
- The underlying HydroSHEDS DEM at 15-arcseconds was found to dramatically over-exaggerate the grosshead of some trial hydropower stations. This was corrected (post search) by including a novel grosshead estimation tool utilising a higher resolution 3-arcseconds DEM with no significant impact on search times. This could be incorporated pre-search in future model trials.
- There are some known weaknesses within the costing model, including the lack of availability of an accurate transmission network dataset and simplification of the road dataset, although the latter had insignificant impact on most project costings.
- It was found that restricting the penstock size to less than 10km gave higher NPV from a catchment as a whole.

- A hydrological model was developed to produce the flow duration curve data necessary based on the G2G hydrological model incorporating strongly performing (in the literature) APHRODITE daily datasets, a daily evapotranspiration dataset developed uniquely for this project (based on the FAO methodology and MERRA data) and the SNOW-17 snow accumulation and ablation model.
- The hydrological model estimated the daily surface water flow volumes across each cell of the YDB DEM at 15-arcseconds resolution from 1979 to 2007, producing the flow duration curves for all rivers with a minimum catchment of $2km^2$.
- Simulating one year of the YDB as whole took approximately 2 weeks, which is slow if trying to produce a long-time series (e.g. 30 years) for catchments at a YDB scale. A compromise was sought here by running each year in parallel, but faster code running in series would produce superior results.
- The genetics based DEOptim evolutionary algorithm was used to optimise the calibration of the hydrological model, comparing the model results to observations across 8 multi-objective efficiency criteria.
- Validation of the final selected parameter set was achieved by comparing observation results compared to model results on a $1,679,569km^2$ sub catchment of the YDB across 2004. This resulted in a mean efficiency criterioa result of 0.97, suggesting the hydrological model performs well for extraction of flow values for use in a hydropower search engine.
- The GLUE methodology was employed on the hydrological model showing uncertainty was greatest during peak floods, but overall the hydrological model was fit-for-purpose.
- The meteorological datasets used were from 1979 to 2007 and showed no significant trend in increasing or decreasing precipitation. However, there was a significant trend in temperatures and evapotranspiration rates (and potentially snowmelt rates) suggesting the potential of reducing energy rates from some hydropower regions. Datasets from beyond 2007 would be required to characterise this further.

8.4 Limitations and further work

An obvious limitation of the work is the choice of resolution. Selecting the HydroSHEDS 15-arcsecond DEM, which dictated the resolution of the model, was a compromise between model accuracy and efficiency. Although the model correlated well with river flow observations, most of the observations were from larger rivers. Hence, it is unclear whether the derived river network is representative of smaller streams high on the catchment. As the width of a cell at 15-arcseconds is approximately 450m, there could be several small streams crossing each grid cell which appear as one in the model. Hence hydropower may appear viable where in fact it may not be. Furthermore, the 15-arcseconds DEM performed poorly in estimating the gross head, which may also be indicative of problems within the hydrological model.

Ideally, the 3-arcseconds DEM would have been used throughout the project, but this would have led to many difficulties. In the current form of the code, the memory requirements and processing time would have been in excess of what was available due the 25x increase in grid cells plus at least 5x decrease in timestep. However, there may be alternative approaches to modelling the Yangtze. If the YDB was broken into hydrologically distinct sub-catchments, the modelling work could take place on smaller areas. Different parameters could be set for each sub-catchment within the hydrological model, rather than requiring one compromise parameter set for the whole basin. This would require adding the outflows from lower order catchments to the inflow of higher order catchments further downstream, a major re-write of the existing code.

It may be possible to perform the calibration work of the hydrological model at a lower resolution (e.g. 15-arcseconds) and then use the best parameter set derived at a higher resolution, easing the burden of calibration (this is the methodology adopted by Duncan (2014) who calibrated at 1km resolution and then applied the best parameter set at 200m resolution). The number of iterations within the hydropower search would also increase dramatically, but the accuracy of the search would be improved. Increased parallelisation would be required to achieve this, alongside replacing some of the R code with C++ code utilising the Rcpp module.

The access road dataset was also simplified due to the density of the data, which took excessive time in finding distances from the scheme to the nearest road. Overall, this only had a small impact on project costs, affecting only some of the more marginal micro-hydro sites.

However, the original dataset could still be used if the modelling work was conducted on smaller sub-catchments as suggested above.

A major limitation was the lack of access to a thorough description of the Chinese transmission network, substituted by a proxy set of habitation. A transmission network GIS dataset covering all voltages was not available within the public domain. To address this would require communication with Chinese organisations, possibly conceding that to produce a thorough model purely from public domain sources is impossible. In fact, access to data in general may be an issue going forward if mapping other catchments in other countries, and will depend on location. Furthermore, the algorithm does not consider loads on the network and assumes all the energy can be transmitted and used.

The meteorological data used is spatially and temporally low resolution (in comparison to the DEM) despite APHRODITE being considered a high-quality product for that part of Asia. Other datasets were rejected as they did not have sufficient historic data for the long time-series required, but it would be interesting to test higher resolution datasets within the hydrological model (e.g. PERSIANN-CCS at approximately 4km resolution and hourly timestep). Hourly data would be more representative of how precipitation and temperature changes diurnally, improving the model accuracy but at the expense of increased data storage and processing time. However, up to date calibration data (i.e. observed river flows) is limited in the public domain for catchments such as the Yangtze, and again may require liaison with in-country organisations to obtain more up to date observations.

Temperature data suggests the mean air temperature of the Yangtze is warming, potentially supporting the hypothesis of climate change. This could impact future resource and hence investigating the impact on hydropower due to climate change, by incorporating a climate model with the hydrological model, would be an important task.

The temporal resolution of the daily evapotranspiration dataset developed for this project is far higher than any other product available. However, like the precipitation and temperature datasets, an hourly evapotranspiration dataset would be even more representative, and the required data is available within the MERRA data suite. However, developing evapotranspiration datasets takes time and it is questionable whether this would add significantly compared to using an off the shelf evapotranspiration product e.g. the evapotranspiration dataset produced by the University of East Anglia Climatic Research Unit,

(2013). This would be an interesting exercise dictating whether the time taken to develop an evapotranspiration dataset with high temporal resolution is worth the effort.

It is difficult to compare the snow model to real observations, particularly as the snow model used needs to maintain a water balance with incoming precipitation. The conceptual index-melt model SNOW-17 adds reality to catchment modelling, but comes at the expense of a significant number of calculations adding to modelling time. It would be a worthy exercise to compare results using a much more simplistic, but faster, degree-day based model.

G2G is a relatively simple hydrological model requiring basic input data requirements. Despite its limited use outside of the UK and being coupled to data from disparate sources, it performed well against observations in calibration. However, G2G does not take into account land use, for example, and other hydrological models could be explored e.g. SWAT.

RETScreen has been used extensively within the literature and the costs generated compare well to already built schemes. However, the equations used in this project date to 2004 and RETScreen was only designed for schemes up to 50MW. Therefore, it would be worthwhile liaising with RETScreen to see if the 2004 equations have been updated in their 2016 release of RETScreen expert. Furthermore, it would be worthwhile exploring with RETScreen the limitations on using the model for schemes beyond 50MW. However, this could be challenging due to the commercialisation of their latest release. A very simple energy pricing method was also used in this work which requires further development.

In this work, the minimum number of turbines was selected for each scheme. Anagnostopoulos and Papantonis (2007) found that having two turbines operating in parallel configured for different design flows improved the economics of a scheme, which could be included in this algorithm.

This modelling work searched for potential ROR schemes without taking into account existing schemes. Only a proportion of the viable schemes suggested would be possible to implement as some would clash with existing dams and reservoirs. Therefore, it is necessary to obtain details of existing dams. Only the natural geography was used in assessing the gross head of potential schemes in this work, but a developer may be interested in the costs and benefits of a higher dam. The associated water storage behind the dam could be calculated by infilling the DEM to the height of a virtual dam. If this was realised then impoundment type projects,

or pumped storage projects, could also be investigated, making the model a powerful tool for the hydropower industry.

Appendix 1 – Contents of the DVD

The accompanying DVD contains two folders entitled Code and Results.

The files within the 'Code' folder are .r files which can be read in Notepad++ (https://notepad-plus-plus.org/) and set up for windows machines. These include:

1. Distributed hydrological model R code.r
2. Hydropower search algorithm R code.r
3. Hydropower conflict algorithm R code.r

The files within the 'Results folder are .csv files which can be read in the R programming language by utilising the following code:

Results <- read.table(X:/folder/resultsname.csv,sep=",",header=TRUE)

where X: and folder represents the drive letter of the location of the results and the resultsname is the title of the results to be read. These include:

1. OriginalsearchIRR5.csv
2. OriginalconflictfreeIRR5.csv
3. OriginalconflictfreeIRR10.csv
4. OriginalconflictfreeIRR15.csv
5. Searchwith3sheadIRR5.csv
6. FinalconflictfreeIRR5.csv
7. FinalconflictfreeIRR10.csv
8. FinalconflictfreeIRR15.csv
9. 100samplescomparinggrosshead

Explanations for the results files are given below:

1. All searches with the original 15-arcseconds estimated gross head with an IRR of ≥5%.

2. The optimised best NPV network after applying the conflict free algorithm on all schemes from results file 1.
3. The optimised best NPV network after applying the conflict free algorithm on all schemes from results file 1 with an IRR ≥ 10%.
4. The optimised best NPV network after applying the conflict free algorithm on all schemes from results file 1 with an IRR ≥ 15%.
5. All the re-assessments of results file 1 after adjusting the gross head to that estimated from the 3-arcseconds DEM.
6. The optimised best NPV network after applying the conflict free algorithm on all schemes from results file 5.
7. The optimised best NPV network after applying the conflict free algorithm on all schemes from results file 5 with an IRR ≥ 10%.
8. The optimised best NPV network after applying the conflict free algorithm on all schemes from results file 1 with an IRR ≥ 15%.
9. Details of the 100 samples taken to compare the estimated 15-arcseconds gross head with the Google Earth estimated gross head and the re-calculated 3-arcseconds gross head method.

Appendix 2 – Published conference paper

Mapping the hydropower resource of the Yangtze River drainage basin

A.D.Walker[1], G. P. Harrison[1], T.Bruce[1], C.A.Greated[2]
[1] Inst. for Energy Systems, University of Edinburgh, King's Buildings, Edinburgh, EH9 3DW

[2] School of Physics and Astronomy, University of Edinburgh, King's Buildings, Edinburgh, EH9 3DW

Abstract

Hydropower is an important component of China's energy mix and this paper summaries work to model and map potential hydropower within the Yangtze drainage basin. To date most of the work has been in development of hydrological modelling over a long time series from 1979 to 2007. Calibration and optimisation of the model on a small sub-catchment of the Yangtze has shown strong correlation between modelled and observed flow duration curves. The outputs from the hydrological modelling will feed into the hydropower search which is still under development. Importantly, the input data sources are publicly available and hence the techniques could be employed on any global catchment.

1 Introduction

Development of low carbon renewable resource is desirable to mitigate future global-warming. Hydropower remains the most important of renewables for electricity production contributing more than the sum of other renewables and nuclear together (BP, 2015). Small scale hydropower, particularly "run of river" is considered one of the most cost-effective and environmentally benign energy technologies and can provide a significant resource and provide power to rural communities off-grid (Paish, 2002).

Due to massive economic growth in China coupled with increasing energy consumption, alternatives to fossil fuels are essential for future sustained growth and to mitigate climate change. China generates almost 25% of global renewable energy but also consumes 23% of world primary energy (BP, 2015*)* emitting almost 25% of all CO_2 emissions (Boden et al.,

2015). China is estimated to be the most resource rich country in the world for hydropower and in 2015 produced 1064 TWh, approximately 19% of China's electricity generation. Mapping of China's hydropower resource would be of use to government (national and local), energy supply companies and overseas agencies interested in renewable energy and/or climate change.

China hosts some of the longest rivers in the world. By length, discharge and catchment area, the Yangtze is by far the biggest of these at 6300km, an average discharge at the river mouth of approximately 30,000m^3s^{-1} and a catchment area of approximately 1.9 million km^2. The upper basin of the Yangtze River includes mountain peaks over 6400m and the mountainous terrain is ideal for hydropower resource. The Three Gorges Dam on the middle reaches of the Yangtze is the largest power station in the world.

This project aims to map the Yangtze basin hydropower resource by creating a mathematical model to estimate long-term hydrographs of river flow employing high resolution datasets of precipitation, temperature and evapotranspiration in conjunction with a suitable digital elevation model (DEM). The results of this model can then be investigated to find economically viable hydropower resource and siting of installations considering both run-of-river and impoundment type schemes. Long term hydrographs of ungauged catchments may also be of use for other purposes (e.g. flood control).

2 Methodology

2.1 Selection and Development of Terrain Datasets

A fundamental aspect of assessing potential hydropower is the need to model the underlying terrain and how the water routes over the terrain as it heads towards the catchment outlet. High quality Digital Elevation Models (DEMs) representing most of the Earth's surface are available as a raster (an array of cells) of heights produced using satellite data. The Shuttle Radar Topography Mission (SRTM) DEM at approximately 90m resolution (Farr et al., 2007) is suitable for derivation of watershed boundaries and drainage networks. However, voids in the SRTM could be a problem in hydrological analysis (Ling et al., 2005) hence the derived HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales) DEM dataset was selected for the model due to the hydrological conditioning (or error reduction) applied (Lehner et al., 2008). The HydroSHEDS dataset improves upon

the SRTM data through a series of automated procedures including void filling, filtering, stream burning and upscaling techniques with manual corrections where necessary available at 3 arc-second, 15 arc-second (approximately 450m) and 30-arc-second grids. Quality assessments indicate the accuracy of the HydroSHEDS dataset significantly exceeds that of other DEMs for creation of watersheds and river maps. Due to the vast scale of the Yangtze basin and computing power available it was decided to model the terrain using the 15 arc-second dataset.

A flow direction raster defines the direction of flow from each cell to its steepest down-slope neighbour derived using the D8 algorithm (O'Callaghan and Mark, 1984). Using standard GIS techniques the slope and aspect of each cell can be determined and the algorithm assumes water follows the steepest path into one of 8 directions. Flow accumulation is derived from the flow direction raster, and is the sum of cells that would ultimately flow into a downstream cell. The HydroSHEDS river lines dataset determines rivers as those with a catchment of 8km^2 or greater, suitable for some basic initialisation of a model but rivers may form from a smaller catchment size within areas of high precipitation, while in dry areas (such as high on the Qinghai Plateau) larger catchment contribution may be necessary before a river forms. The boundary of the Yangtze drainage basin was determined within ArcGIS using the HydroSHEDS Asia flow accumulation raster by identifying the drainage point (mouth) of the Yangtze River. This enabled a suite of Yangtze-specific datasets including DEM, flow direction, flow accumulation and river network rasters (see Figure 1).

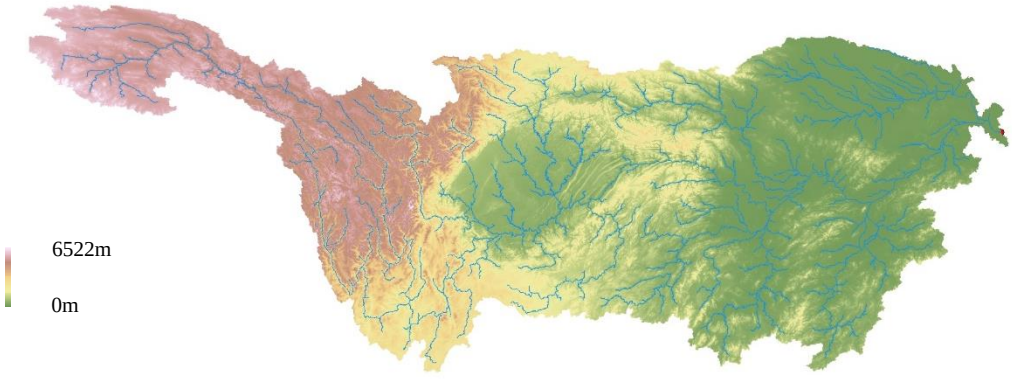


Figure 1: DEM of Yangtze basin with rivers overlaid (large rivers only) (produced using ArcGIS)

Further input datasets determine the starting conditions for the model. A model starting dry could take a significant amount of time to "warm-up" and hence the starting in-soil/ground storage, baseflow and snow conditions were set from the Modern-Era Retrospective Analysis for Research and Applications (MERRA) data suite (Rienecker et al., 2011). For starting overland runoff conditions, the flow accumulation raster was 'weighted' using average long-term overland flow from the Global Runoff Data Centre (GRDC) Composite Runoff Fields dataset produced by the Bundesantalt für Gewasserkunde (BfG) (Fekete et al., 2002). These starting condition datasets are coarse and most likely an over or under estimate of actual conditions but it was felt that this would enable the model to "warm up" faster than if starting conditions were assumed to be completely dry and particularly as it would ensure rivers are flowing from the start. In practice, the model is also ran for a few months in advance of the actual starting month to enable the model to warm-up.

2.2 Selection and Development of Meteorological Datasets

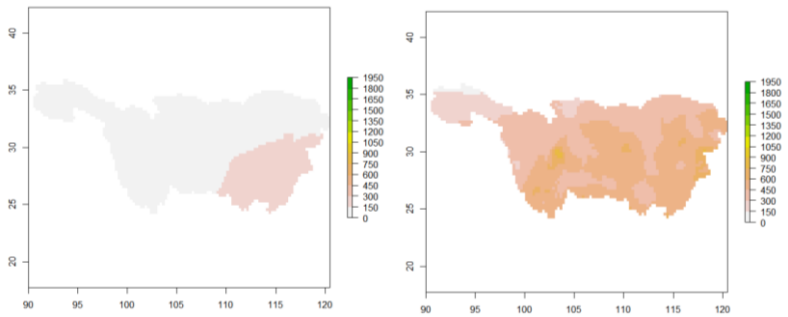
Precipitation and temperature data is another key input to hydrological models and there are a number of gridded meteorological datasets available at various temporal and spatial resolutions covering a significant proportion of the globe (Schneider, 2013). Ideally the data would be both of high spatial and temporal resolution (sub-daily) but most of the sub-daily datasets are limited in their history or global coverage. Hydropower estimation requires long time-series of data. The MERRA data suite holds gridded precipitation and temperature data dating back to 1979 at hourly resolution but is of low spatial resolution. Therefore the Asian Precipitation Highly Resolved Observational Data Integration Towards Evaluation of Water Resources (APHRODITE) precipitation dataset (Yatagai et al., 2012) and temperature dataset (Yasutomi et al., 2011) were selected due to their relatively high spatial resolution (0.25 by 0.25 degrees, ~25km), daily time step and availability from 1973 to 2007 (1950 for the precipitation dataset).

Precipitation data determines the input of moisture into the drainage basin whereas temperature determines whether precipitation falls as snow or rain and also contributes to evapotranspiration rates. The APHRODITE precipitation dataset was created using data from between 5,000 and 12,000 rain gauge stations across Asia and has significantly improved rainfall data for regions such as the Himalayas. The temperature dataset was created from a significant number of observation stations, up to 3 times the number within the Global

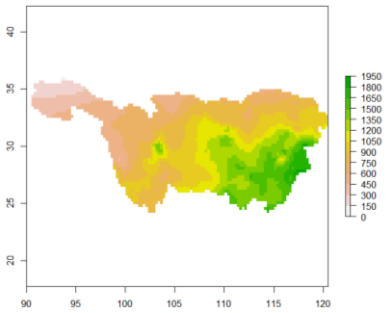
Telecommunication System used for the creation of most other datasets of this nature. Temperature is compensated by the standard adiabatic lapse rate due to the difference in height of the DEM and the average heights of the temperature dataset.

Evapotranspiration accounts for loss of water due to evaporation from land and water bodies to the atmosphere together with plant transpiration (movement of water within plants and loss of vapour from leaf stomata). A daily dataset from 1979 to 2007 was created (by this project) using MERRA data employing the Food and Agriculture Organisation (FAO56) method based on the Penman-Monteith equation (Allen et al., 2006). Contributing MERRA datasets include daily maximum and minimum temperatures, daily maximum and minimum specific humidity at 2m above displacement height, mean air pressure, northerly and easterly windspeed, leaf area index, displacement height, net surface downward shortwave radiation flux, albedo fraction and emitted and absorbed longwave radiation at the surface.

Note that to date there has been no attempt made to account for diurnal changes in meteorological conditions within the model and assumes temperature is constant throughout each day and hourly precipitation and evapotranspiration is $1/24^{th}$ of the daily amount. Meteorological datasets with a higher temporal resolution (i.e. sub daily) would improve the model. The plots in Figures 2 and 3 show the annual, summer and winter average precipitation and temperature for 1979 to 2007. Summers in the Yangtze are categorized by high levels of precipitation and warm temperatures whereas winters are cool and dry, particularly so in the western portion of the catchment.

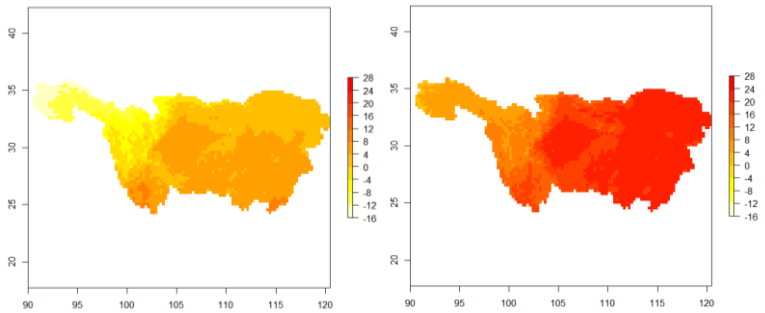


Winter mean 82mm *Summer mean 432mm*



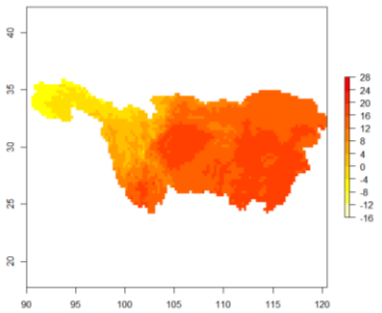
Annual mean 954mm

Figure 2: Winter (Jan-Mar), summer (Jul-Sep) and annual mean precipitation in the Yangtze drainage basin 1979 – 2007 using APHRODITE data (Yatagai et al., 2012)



Winter mean 1.6°C

Summer mean 17.2°C



Annual mean 11.7°C

Figure 3: Winter (Jan-Mar), summer (Jul-Sep) and annual mean temperature in the Yangtze drainage basin 1979 – 2007 using APHRODITE data (Yasutomi et al., 2011)

Although only over a relatively short period of time (1979-2007), analysis of the rolling 12-month average precipitation and temperature shows a slight decreasing trend in precipitation and a pronounced increasing trend in temperature (*see figure 4*) – possibly due to the impact of climate change. This may result in future increased evapotranspiration, and a change in the snow regime, which could have a long-term impact on the hydropower resource.

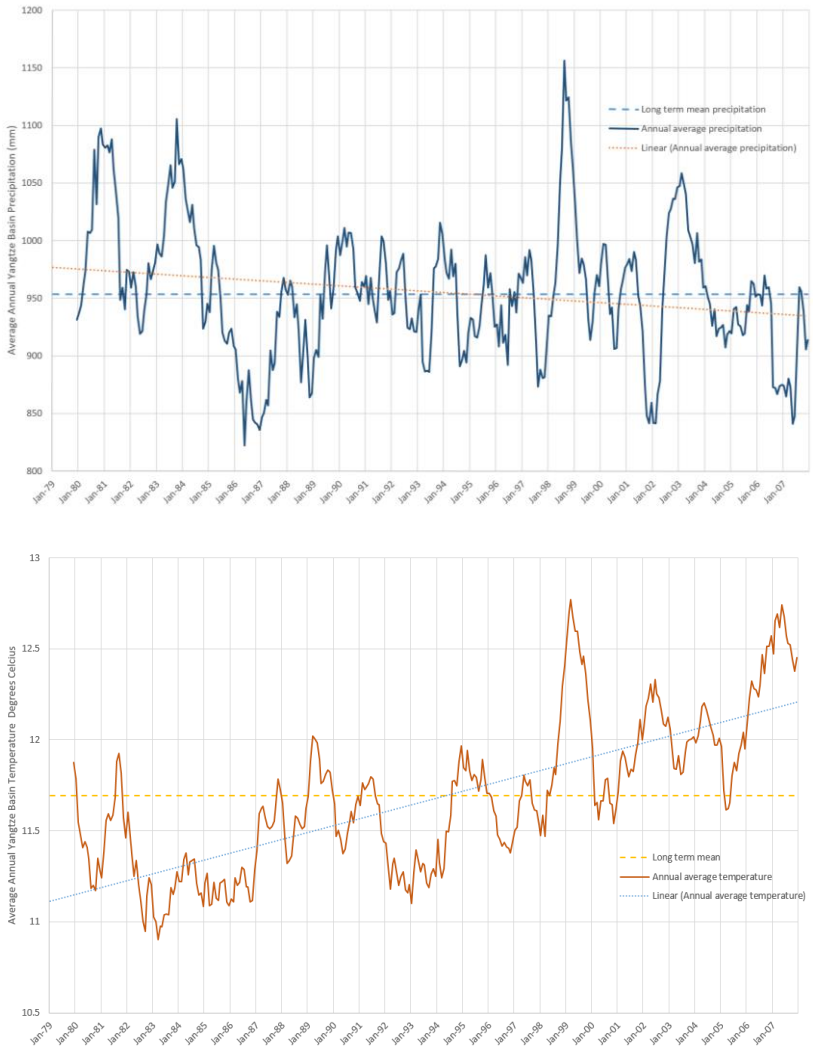


Figure 4: Trends in average annual precipitation (top) and temperature (below) in the Yangtze basin between 1979- 2007 using APHRODITE data (Yatagai et al., 2012; Yasutomi et al., 2011)

2.3 Inclusion of a Snowmelt Model

Temperature index models or degree day snowmelt models are often employed within runoff models, assuming a relationship between air temperature and melt when above a threshold temperature: the degree day factor. However, this melt factor is not constant and can vary spatially and seasonally and hence there has been a gradual transition towards energy-balance models (Hock, 2003). This hydrological model employs the widely quoted National Weather Service River Forecast/Snow Accumulation and Ablation Snow-17 model (Anderson, 1973; Anderson 2006). Snow-17 is a conceptual model incorporating most of the physical processes that take place within snow cover but only in a simplified form. Despite its sophistication there

is no data to calibrate this snow model against and therefore standard initialisation parameters have been used. Furthermore, it is computationally time expensive when running in catchments the size of the Yangtze and therefore a simpler model may be considered in the future.

2.4 An overview of the grid based hydrological model

Building on a study of Scotland's hydropower resource (Duncan, 2014), the distributed Grid-to-Grid (G2G) model (Bell et al., 2007) was selected to simulate how water flows across the terrain due to its ease of integration with grid based datasets and proven ability to simulate flows at high resolution, although admittedly only in the UK thus far. The equations employed in the G2G model are used to determine the runoff production and routing algorithms of this model, built using R code integrating the datasets previously discussed, with some of the flow routing coded in C++, and versions built for both Linux and Windows.

After some initialisation and loading of input datasets, daily meteorological files are loaded into the model at the start of each day's simulation, spatially modified to the model's resolution discussed and the precipitation and potential evapotranspiration found. Within each grid square a water balance is maintained with each cell having a finite store related to the topographic gradient (shallow gradients permit greater storage capacity), with excess water forming runoff (a fast surface kinematic wave). Drainage from the store adds to baseflow (a slow subsurface kinematic wave) and return flow links the baseflow to the surface runoff. Routing of the runoff is based on the kinematic wave approximation of the St. Venant equations for gradually-varying flow in open channels. An additional layer of complexity is added to simulate real hill slope conditions by integrating the Probability Distributed Soil-Moisture formulation (Moore, 1985) and assumes a certain proportion of the grid square is saturated and generating runoff even when rainfall amounts are small – otherwise the whole grid square would have to be saturated before runoff was produced. The surface/subsurface wave speeds and return flow rates (which can differ for land and river), the drainage rate and maximum cell storage form a set of parameters that require calibration and optimisation.

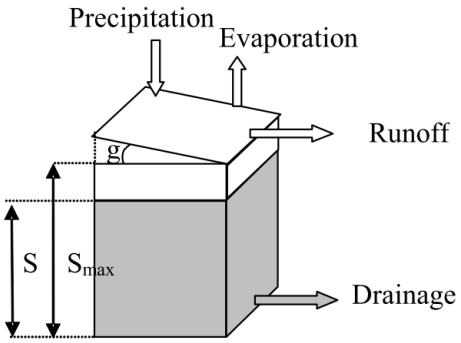


Figure 5: Overview of the G2G model grid-box storage illustrating components of the water balance (Bell et al., 2007)

2.5 Running the code, calibration and outputs

The code can be run over any time period between January 1979 and December 2007 and with a specific set of parameters or in calibration mode employing Differential Evolution (Storn and Price, 1997). Differential evolution optimises the objective function comparing results from the simulated flow at the catchment outlet to observed flows by sampling a wide parameter space. New candidate solutions are combined with existing solutions keeping solutions with the best score or fitness. A number of objective functions for hydrological model assessment (Krause et al., 2005) are built into the model with only one objective function selected at any one time (typically Nash Sutcliffe efficiency). As timing issues with respect to the input data could reduce objective function results when comparing daily and simulated daily discharge, the catchment flow duration curves (FDC) are compared in the calibration. Ultimately, it is the flow duration curve that is important in hydropower assessment.

Calibration data is publicly available for a small number of rivers of various catchment sizes within the Yangtze from the GRDC river discharge data (GRDC, 2015), although there are only 12 such points in the Yangtze catchment which is nearly 2 million km^2. Calibration requires hundreds/thousands of model runs trialling different parameters within the model space but large catchments require significant computing time for each run. Hence, calibration is carried out on smaller catchments with the premise that these calibration parameters will hold for larger catchments. To speed up the calibration process the code is run in parallel on the University of Edinburgh EDCF Linux Compute Cluster. Once calibrated, the model is run with an optimal set of parameters with the hourly and daily flow recorded across the catchment

and saved as daily raster files from which the modelled FDC percentile flow is obtained. The river network and the flow duration rasters feed into the hydropower search.

2.6 Testing the model

Testing of the model was carried out on the catchment above the Laoguan He (Laoguan River) Xixia monitoring station – a 3205km^2 catchment (*see figure 6)*. The catchment is on the northern boundary of the Yangtze drainage basin with the drainage point approximately 265km from the Yangtze itself and 840km from the mouth. Calibration was carried out over 12 months starting in January 1979 with two warm-up months, with the best performing parameter sets ran again from 1979 to 1992 (the full length of the observation data). The best performing parameter set was then selected to run from 1979 to 2007.

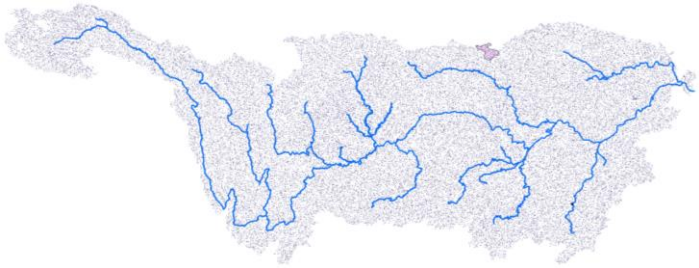


Figure 6: Location of the Laoguan He catchment for testing the model (pink shade)

2.7 Hydropower search

The hydropower search is still in development. The proposed methodology is to place points at intervals across the river network found through hydrological modelling and extend a 'virtual penstock' to points downstream. By trialling various penstock diameters, and design flows, and employing the flow duration curve percentile rasters, a number of iterations can be costed. Economically viable configurations will be extracted and mapped.

3 Results

Over a two-week period, 1588 model runs for the year 1980 were performed on the Laoguan He Xixia catchment. Results with a total modelled runoff volume of between 75% and 125% of the observed runoff volume at the catchment outlet were retained for further analysis. These were sorted by each of the objective functions and the top 5 performing parameter sets for each again kept for further analysis, resulting in approximately 40 parameter sets. The FDC curves and daily flow plots were inspected visually and the best 5 selected for a longer

calibration run (1979 to 1992). Of the 11 objective functions tested the results ranged from 0.715 to 0.993, and 8 of the 11 had a result greater than 0.95. Only the modified form of the Nash Sutcliffe efficiency and the Nash Sutcliffe efficiency using logarithmic values had a result less than 0.9 (0.822 and 0.715 respectively), which are both designed to increase the sensitivity of low flow values. Once again, the best performing parameter set during the longer calibration was selected for the full length model run (1979 to 2007).

Example Screenshots during running of the model are shown overleaf in *Figure 7*:

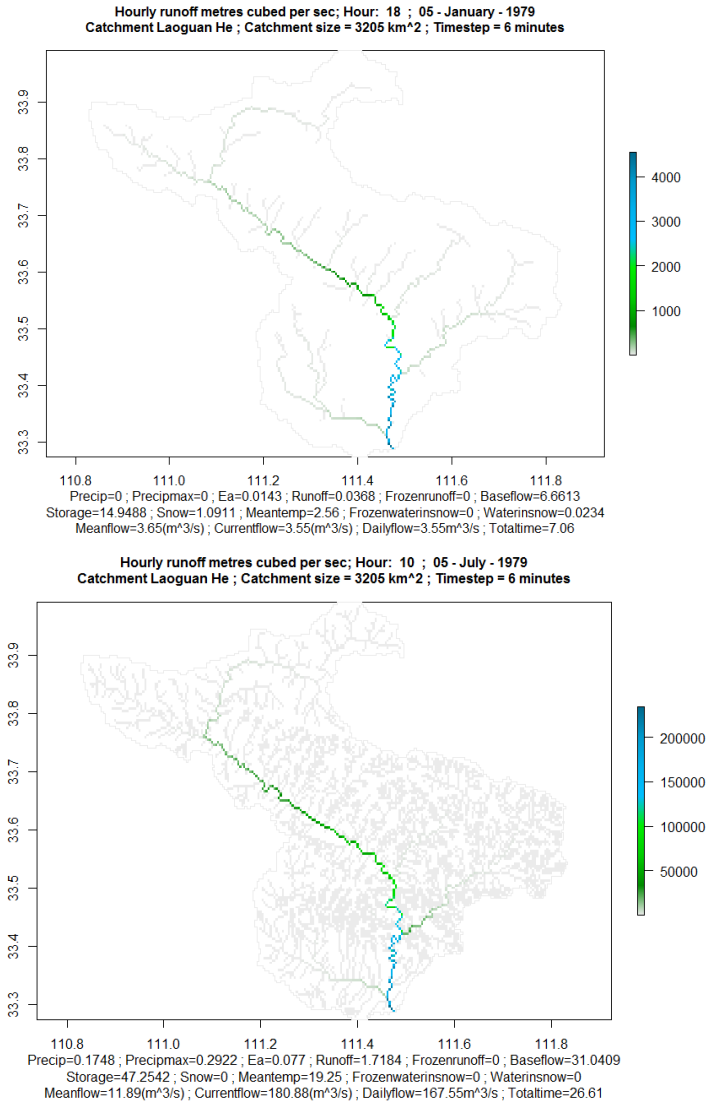


Figure 7: Example screenshots during running of the model showing surface runoff (l/s) (only > 5l/s shown) in January (left) and July (right)
(Values below the plot are catchment averages in mm except for temperature (°C), time (mins) and flow (m^3s^{-1}))

At the end of a calibration run a daily flow plot and an FDC curve are generated showing the modelled data in red and the observed data in green. *Figure 8* (1980 only) and f*igure 9* (1979 to 1992) show the modelled and observed data for daily flow and the FDC.

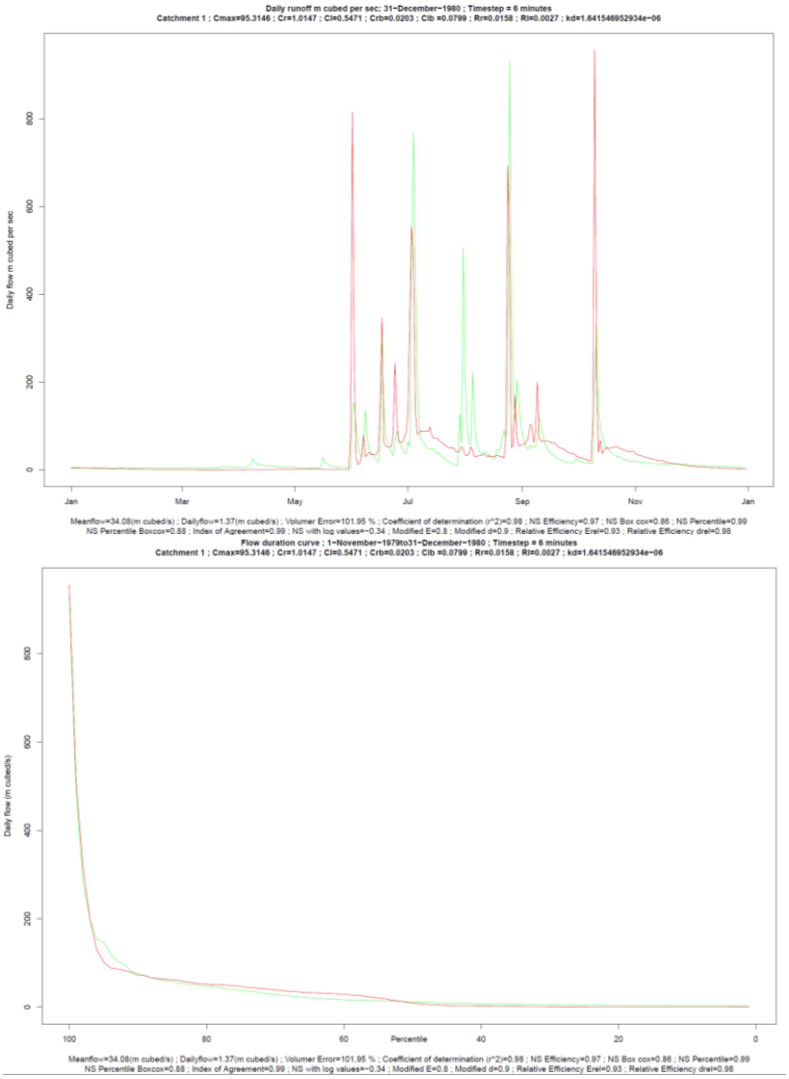


Figure 8: (Left) Modelled (red) and observed (green) daily surface runoff and (Right) modelled (red) and observed (green) flow duration curve – both for 1980 using the optimised parameter set at the Laoguan He Xixia catchment outlet

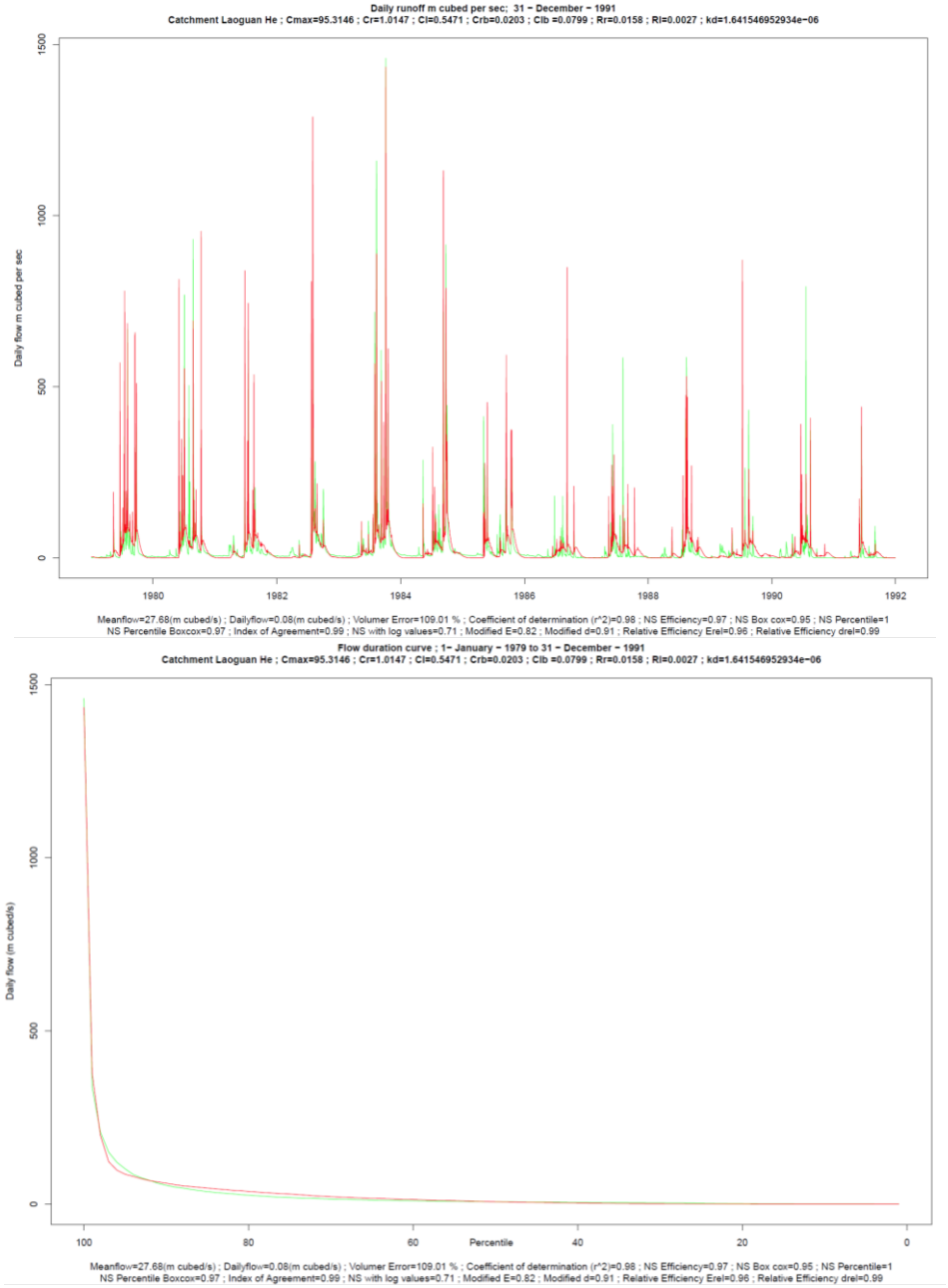


Figure 9: (Left) Modelled (red) and observed (green) daily surface runoff and (Right) modelled (red) and observed (green) flow duration curve – both for 1980 using the optimised parameter set at the Laoguan He Xixia catchment outlet

The 1979-2007 daily flow and FDC curves using the optimised parameter set for the Laoguan He Xixia catchment is shown below in *Figure 10*:

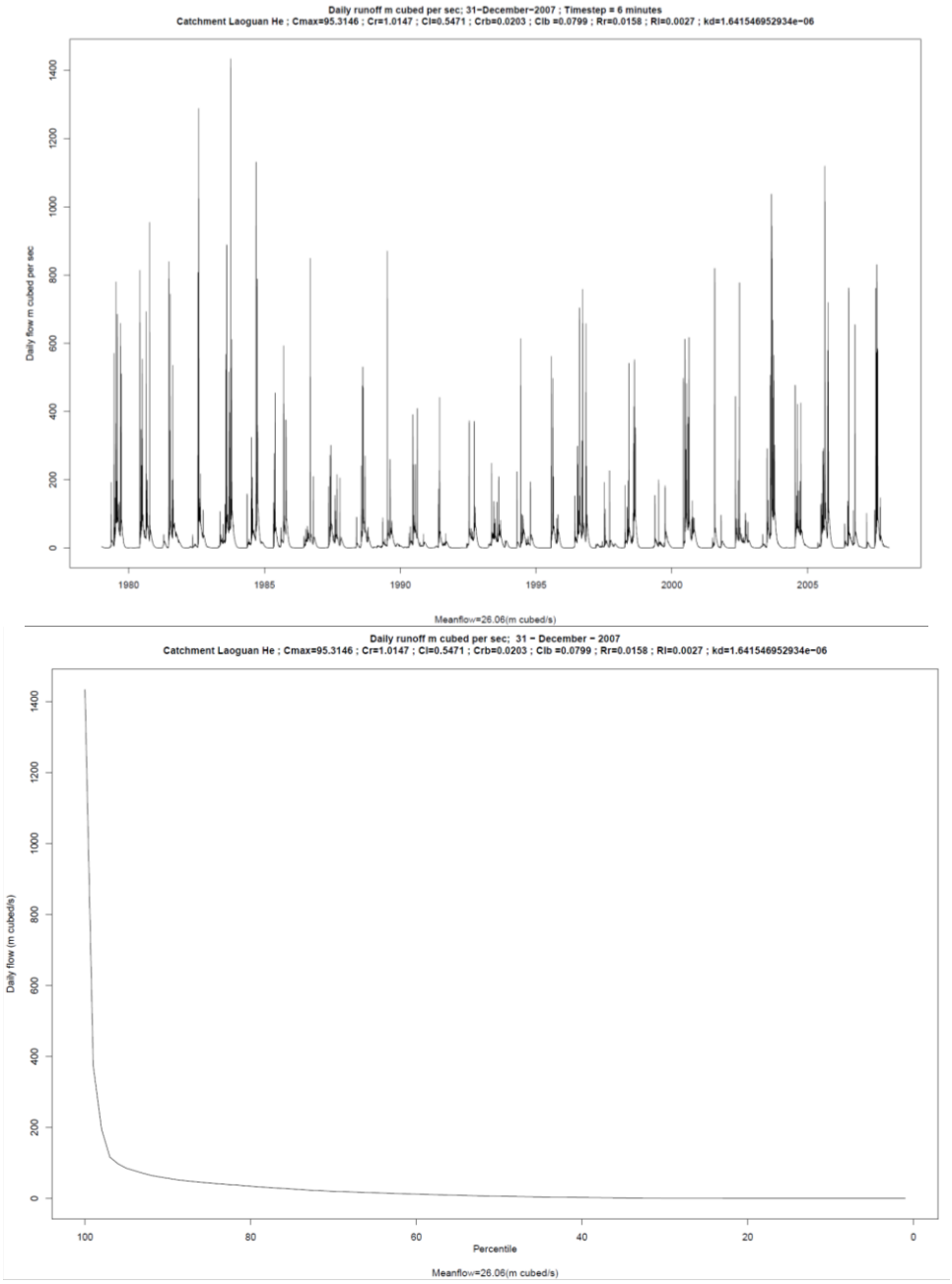


Figure 10: (Left) Screen plot of the full 1979 to 2007 daily surface runoff (modelled data only) and (Right) screen plot of the full 1979 to 2007 flow duration curve (modelled data only) for the Laoguan He Xixia catchment using the optimised parameter set – units m^3s^{-1}

Combing the daily flow enables the percentile values at each point in the catchment to be extracted: examples for 100th, 50th and 5th percentile plots (or Q0, Q50, Q95) are shown in *Figure 11:*

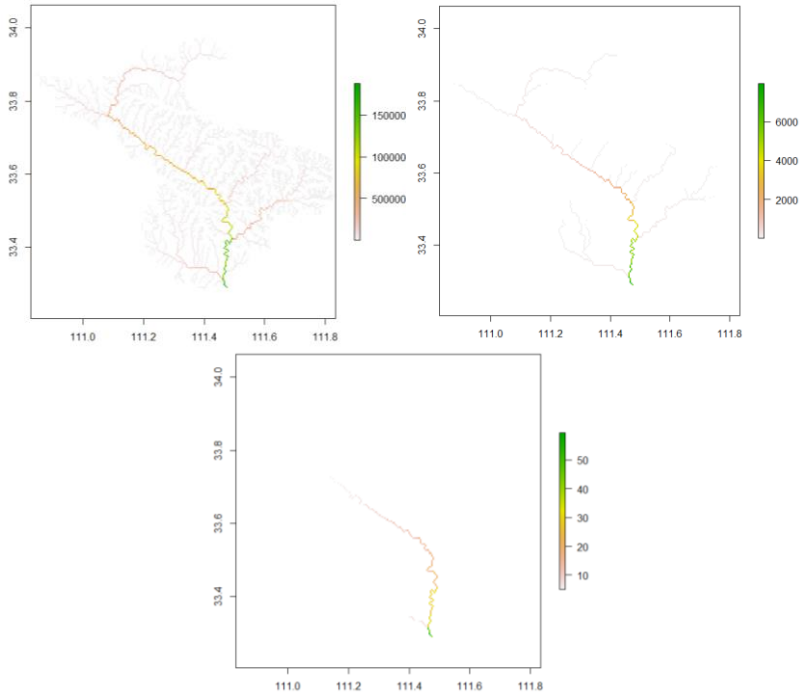


Figure 11: Modelled 100th percentile (Q0) (left), 50th percentile (Q50) (middle) and 5th percentile (Q95) (right) surface runoff data for the Laoguan He Xixia catchment (3205km^2) from 1979 to 2007 (for flow >5l/s) (units l/s)

4 Conclusions

Renewable technologies are of ever increasing importance due to both energy requirements and climate change, particularly in countries such as China where there has been massive economic growth. Mapping of hydropower in basins such as the Yangtze would be useful to agencies internal and external to China, and the techniques employed transferable to other catchments around the world as the input data is global and accessible. The model developed has been shown to correlate well compared to observed data with high-performing results across a number of objective function types, particularly relating to the FDC. However, there are errors and simplification within both the input and observed data and these need to be examined more closely. A strategy needs to be developed on how the results from small sub-catchments, such as Laoguan He Xixia, can be used to model the hydrology of the whole Yangtze catchment as computing resources are limited. The resulting percentile flow rasters are easily integrated into a hydropower search methodology.

References

Allen, R.G., W.O. Pruitt, J.L. Wright, T.A. Howell, F. Ventura, R. Snyder, D. Itenfisu, P. Steduto, J. Berengena, J.B. Yrisarry, M. Smith (2006), A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO56 Penman-Monteith method. *Agricultural Water Management*, **81**(1): 1-22.

Anderson, E.A. (1973), SNOW–17 a model based on temperature and precipitation. *National Oceanic and Atmospheric Administration (NOAA)*, publication SW-102, Washington DC, USA.

Anderson, E.A. (2006), Snow Accumulation and Ablation Model – SNOW-17. *National Oceanic and Atmospheric Administration (NOAA) / National Weather Service*, Washington DC, USA.

Bell, V.A., A. Kay, R. Jones, R. Moore (2007), Development of a high resolution grid- based river flow model for use with regional climate model output. *Hydrology and Earth System Sciences*, **11**(1):532–549.

Boden, T.A., G. Marland, R.J. Andres (2015), Global, Regional, and National Fossil-Fuel CO_2 Emissions. *Carbon Dioxide Information Analysis Center*, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. doi 10.3334/CDIAC/00001_V2015

British Petroleum (2015), Statistical Review of World Energy 2015. British Petroleum, London, UK.

Duncan, N. (2012). Mapping the Hydropower Resource of Scotland. Ph.D. Thesis. Edinburgh University, UK.

Farr, T.G, P.A. Rosen, E. Caro, R. Crippen., R. Duren; S. Hensley; M. Kobrick, M.Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffe, J.Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, D. Alsdorf (2007), The Shuttle Radar Topography Mission. *Reviews of Geophysics* **45** (2): RG2004. doi:10.1029/2005RG000183

Fekete, B., C. Vörösmarty, W. Grabs (2002): Global composite runoff fields on observed river discharge and simulated water balances. Water System Analysis Group, University of New Hampshire, and Global Runoff Data Centre. Koblenz, Germany: Federal Institute of Hydrology (BfG).

GRDC (2015), River Discharge Data. The Global Runoff Data Centre, Federal Institute of Hydrology (BfG), 56068 Koblenz, Germany

Hock, R. (2003), Temperature index melt modelling in mountain areas. *Journal of Hydrology* **282**(1): 104-115.

Krause, P., D. P. Boyle, and F. Bäse (2005), Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* **5**: 89-97.

Lehner, B., K. Verdin, A. Jarvis (2008), Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales - New global hydrography derived from spaceborne elevation data. *Eos, Transactions, AGU*, **89**(10): 93-94.

Ling, F., Q. W. Zhang, and C. Wang (2007), Filling voids of SRTM with Landsat sensor imagery in rugged terrain. *International Journal of Remote Sensing* **28**(2): 465-471.

O'Callaghan, J.F. and D.M. Mark (1984), The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*, **28**(3):323–344. ISSN 0734189X. doi:10.1016/S0734-189X(84)80011-0.

Paish, O. (2002). Small hydro power: technology and current status. *Renewable and sustainable energy reviews*, **6**(6): 537-556.

Rienecker, M.M., M.J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M.G. Bosilovich, S.D. Schubert, L. Takacs, G.-K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, et al. (2011). *MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications*. *J. Climate*, **24**: 3624-3648, doi:10.1175/JCLI-D-11-00015.1

Schneider D.P., C. Deser, J. Fasullo, K.E. Trenberth (2013), Climate Data Guide Spurs Discovery and Understanding. Earth and Space Science News, **94** (13): 121-122.

Storn, R. and K. Price (1997). Differential Evolution – A simple and effective heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, **11:**341-359.

Yasutomi, N., A. Hamada, A. Yatagai (2011), Development of a long-term daily gridded temperature dataset and its application to rain/snow discrimination of daily precipitation, *Global Environmental Research*, **V15N2**:165-172.

Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi, A. Kitoh (2012), APHRODITE: Constructing a Long-term Daily Gridded Precipitation Dataset for Asia based on a Dense Network of Rain Gauges. *Bulletin of American Meteorological Society*, doi:10.1175/BAMS-D-11-00122.1.

Authors

Antony Walker is a PhD student within the Institute of Energy Systems at the University of Edinburgh researching the subject “Mapping the Hydropower Resource of the Yangtze Drainage basin”. Antony has a BEng in Mechanical Engineering from the University of Strathclyde and an MSc in Earth Sciences (climate science) from the Open University. For many years Antony worked in manufacturing (electronics and food) as a process engineer before joining Atkins’ nuclear energy department. Whilst living overseas in China (and earlier India) Antony was inspired to start a PhD relating to renewable technologies and China.

Professor Gareth Harrison is the Bert Whittington Chair at the Institute of Energy Systems at the University of Edinburgh. Gareth’s PhD was in Electrical Power Engineering, 2001, following a BEng in Electrical and Mechanical Engineering, both at the University of Edinburgh. He is on the Editorial board member of the International Journal of Emerging Electric Power Systems and on the International Advisory Board of Wiley Interdisciplinary Reviews: Energy and Environment. Gareth is a chartered engineer, a senior member of the institute of electrical and electronics engineers and a member of the institute of electrical engineers. Gareth’s specialities include renewable resource assessment, network integration of renewable energy, multi-vector energy systems/energy systems integration, climate change impacts on renewable energy and life cycle assessment (or carbon footprints).

Dr Tom Bruce is a senior lecturer within the Institute of Energy Systems, University of Edinburgh and Director of Internationalisation. Tom’s Ph.D. subject was "Violent wave action at seawalls and breakwaters", 2006 following a BSc in Astrophysics and MSc in Astronomical Technologies, all studied at the University of Edinburgh. Tom is a co-manager of the EPSRC Coastal Structures Network and a member of the International Society of Offshore and Polar Engineers (ISOPE). His specialities are wave - structure interaction, wave hydrodynamics and flow measurement (Particle Image Velocimetry (PIV)).

Professor Clive Greated is a senior honorary professorial fellow within the School of Physics and Astronomy at the University of Edinburgh and has many years of experience researching ocean wave behaviour, fluid mechanics and acoustics. Following his PhD, Clive has lectured and lead research groups at the University of Cambridge, University of Southampton, the Danish Institute of Hydraulics and over 40 years at the University of Edinburgh as a professor of fluid mechanics.

Bibliography

Aggidis, G.A., Luchinskaya, E., Rothschild, R. and Howard, D.C. (2010). The costs of small-scale hydro power production: Impact on the development of existing potential. *Renewable Energy*, **35**(12):2632-2638. doi: 10.1016/j.renene.2010.04.008

Agrawala, S., Raksakulthai, V., van Aalst, M., Larsen, P., Smith, J. and Reynolds, J. (2003). Development and climate change in Nepal: Focus on water resources and hydropower (pp. 14-28). Paris: OECD

Agterberg, F.P. (1964). Statistical techniques for geological data. *Tectonophysics*, **1**(3):233-255. doi: 10.1016/0040-1951(64)90006-X

Airbus Defence and Space (2016). SPOT Satellite Imagery. Airbus Defence and Space, Leiden, The Netherlands [Available online at http://www.intelligence-airbusds.com/en/143-spot-satellite-imagery]

Airbus Defence and Space (2016). WorldDEM™. Airbus Defence and Space, Leiden, The Netherlands [Available online at http://www.intelligence-airbusds.com/worlddem/]

Alavian, V., Qaddumi, H.M., Dickson, E., Diez, S.M., Danilenko, A.V. Hirji, R.F., Puz, G., Pizarro, C., Jacobsen, M. and Blankespoor, B. (2009). Water and Climate Change: Understanding the Risks and Making Climate-Smart Investment Decisions. Final Report. World Bank: Washington, D.C.

Alexander, K.V. and Giddens, E.P. (2008). Optimum penstocks for low head microhydro schemes. *Renewable Energy*, **33**(3):507-519. doi: 10.1016/j.renene.2007.01.009

Allen, R.G. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Technical Report 56, Food and Agriculture Organization of the United Nations.

Alterach, J., Peviani, M., Davitti, A., Vergata, M., Ciaccia, G. and Fontini, F. (2009). Evaluation of the remaining hydro potential in Italy. *The international journal on hydropower & dams*, **16**(5), p.56.

Alternative Energy Tutorials (2018). Water Turbine Design. Alternative Energy Tutorials [Available online at: http://www.alternative-energy-tutorials.com/hydro-energy/water-turbine-design.html]

Anagnostopoulos, J.S. and Papantonis, D.E. (2007). Optimal sizing of a run-of-river small hydropower plant. *Energy Conversion and Management*, **48**(10):2663-2670. doi: 10.1016/j.enconman.2007.04.016

Anderson, E.A. (1973). SNOW–17 a model based on temperature and precipitation. National Oceanic and Atmospheric Administration (NOAA), publication SW-102, Washington DC, USA

Anderson, E.A. (2006). Snow Accumulation and Ablation Model – SNOW-17. National Oceanic and Atmospheric Administration (NOAA) / National Weather Service, Washington DC, USA

Anderson, D., Moggridge, H., Warren, P. and Shucksmith, J. (2015). The impacts of 'run-of-river'hydropower on the physical and ecological condition of rivers. *Water and Environment Journal*, **29**(2):268-276. doi: 10.1111/wej.12101

Andreadis, K.M., Schumann, G.J.P. and Pavelsky, T. (2013). A simple global river bankfull width and depth database. *Water Resources Research*, **49**(10):7164-7168

Ansar, A., Flyvbjerg, B., Budzier, A. and Lunn, D. (2014). Should we build more large dams? The actual costs of hydropower megaproject development. *Energy Policy*, **69**:43-56. doi: 10.1016/j.enpol.2013.10.069

Ardia, D., Mullen, K.M., Peterson, B.G. , Ulrich, J., Boudt, K. (2016). 'DEoptim': Differential Evolution in 'R'. version 2.2-4

Arnold, J.G., Srinivasan, R., Muttiah, R.S. and Williams, J.R. (1998). Large area hydrologic modelling and assessment part I: model development. JAWRA *Journal of the American Water Resources Association*, **34**(1):73-89. doi:10.1111/j.1752-1688.1998.tb05961.x

Ashouri, H., Hsu, K.L., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D., Nelson, B.R. and Prat, O.P. (2015). PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, **96**(1), pp.69-83. doi: 10.1175/BAMS-D-13-00068.1

Baldocchi, D. (2014). Measuring fluxes of trace gases and energy between ecosystems and the atmosphere–the state and future of the eddy covariance method. *Global change biology*, **20**(12):3600-3609. doi: 10.1111/gcb.12649

Ballance, A., Stephenson, D., Chapman, R.A. and Muller, J. (2000). A geographic information systems analysis of hydro power potential in South Africa. *Journal of Hydroinformatics*, **2**(4):247-254.

Bell, V.A. and Moore, R.J. (1998). A grid-based distributed flood forecasting model for use with weather radar data: Part 1. Formulation. *Hydrology and Earth System Sciences Discussions*, **2**(2/3):265-281

Bell, V.A., Kay, A.L., Jones, R.G. and Moore, R.J. (2007a). Use of a grid-based hydrological model and regional climate model outputs to assess changing flood risk. *International Journal of Climatology*, **27**(12):1657-1671. doi: 10.1002/joc.1539

Bell, V.A., Kay, A.L., Jones, R.G. and Moore, R.J. (2007b). Development of a high-resolution grid-based river flow model for use with regional climate model output. *Hydrology and Earth System Sciences*, **11**(1):532-549. doi: 10.5194/hess-11-532-2007

Beven, K. and Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological processes*, **6**(3):279-298. doi: 10.1002/hyp.3360060305

Beven, K. (2001a). On hypothesis testing in hydrology. *Hydrological Processes*, **15**(9):1655–1657. ISSN 0885-6087. doi:10.1002/hyp.436.

Beven, K. (2012). Rainfall - Runoff Modelling: The Primer. Wiley-Blackwell. ISBN 978-0-470-71459-1

Beygelzimer, A., Kakadet, S., Langford J., Arya, S., Mount, D., Li, S (2013). FNN: Fast Nearest Neighbour Search Algorithms and Applications v 1.1

Bivand, R., Rundel, C., Pebesma, E., Stuetz, R., Hufthammer, K.O. (2017a). rgeos: Interface to Geometry Engine - Open Source (GEOS) v0.3-23

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., Rouault, E., (2015). rgdal: Bindings for the Geospatial Data Abstraction Library v1.2-10

Bivand, R., Lewin-Koh, N., Pebesma, E., Archer, E., Baddeley, A., Bearman, N., Bibiko, H-J., Brey, S., Callahan, J., Carrillo, G., Dray, S., Forrest, D.,, Friendly, M., Giraudoux, P., Golicher, D., Gómez Rubio, V., Hausmann, P., Hufthammer, K.O., Jagger, T., Johnson, K., Luque, S., MacQueen, D., Niccolai, A., Lamigueiro, O.P., Short, T., Snow, G., Stabler, B., Stokely, M., Turner, R. (2017b). maptools: Tools for Reading and Handling Spatial Objects v 0.9.2

Bliss, A., Hock, R. and Cogley, J.G. (2013). A new inventory of mountain glaciers and ice caps for the Antarctic periphery. *Annals of Glaciology*, **54**(63):191-199. doi:10.3189/2013AoG63A377

Bond, T.C., Doherty, S.J., Fahey, D.W., Forster, P.M., Berntsen, T., DeAngelo, B.J., Flanner, M.G., Ghan, S., Kärcher, B., Koch, D. and Kinne, S. (2013). Bounding the role of black carbon in the climate system: A scientific assessment. *Journal of Geophysical Research: Atmospheres*, **118**(11):5380-5552. doi: 10.1002/jgrd.5017

Boyle, G. (2004). Renewable energy: power for a sustainable future (Vol. 2) – second edition. OXFORD university press

Boyle, D.P., Gupta, H.V. and Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, **36**(12):3663-3674. doi:10.1029/2000WR900207

Box, G.E.P. and Cox, D.R. (1981). An Analysis of Transformations Revisited, Rebutted.

BP (2017a). BP Statistical Review of World Energy, BP p.l.c, London SW1Y 4PD UK

BP (2017b). BP Energy Outlook, BP p.l.c, London SW1Y 4PD UK

Brauer, M., Freedman, G., Frostad, J., Van Donkelaar, A., Martin, R.V., Dentener, F., Dingenen, R.V., Estep, K., Amini, H., Apte, J.S. and Balakrishnan, K. (2015). Ambient air pollution exposure estimation for the global burden of disease 2013. *Environmental science & technology*, **50**(1):79-88

Britannica (2017). Yangtze River. Encyclopaedia Britannica [Available online at https://www.britannica.com/place/Yangtze-River]

Brown, R.D. and Brasnett, B. (2010). Canadian Meteorological Centre (CMC) daily snow depth analysis data. Environment Canada.

Burba, G. (2013). Eddy covariance method for scientific, industrial, agricultural and regulatory applications: A field book on measuring ecosystem gas exchange and areal emission rates. *LI-Cor Biosciences*. doi: 10.13140/RG.2.1.4247.8561

Burrough, P.A., McDonnell (1998). Principles of geographical information systems. *Economic Geography*, **75**(4):.422.

Cactus (2001). Cactus 2000 Converter for humidity of air. [Available online at https://www.cactus2000.de/uk/unit/masshum.shtml]

Cammalleri, C., Anderson, M.C., Gao, F., Hain, C.R. and Kustas, W.P. (2014). Mapping daily evapotranspiration at field scales over rainfed and irrigated agricultural areas using remote sensing data fusion. *Agricultural and forest meteorology*, **186**:1-11. doi: 10.1016/j.agrformet.2013.11.001

Ceglar, A., Toreti, A., Balsamo, G. and Kobayashi, S. (2017). Precipitation over Monsoon Asia: A Comparison of Reanalyses and Observations. *Journal of Climate*, **30**(2):465-476. doi: 10.1175/JCLI-D-16-0227.1

Centre for Geographic Analysis (2015). China GIS Data. Harvard University, Boston, USA [Available online at http://gis.harvard.edu/resources/data/china-gis-data]

CGIAR (2008). SRTM 90m Digital Elevation Database v4.1. Consultative Group for International Agricultural Research, Montpellier, France [Available online at http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1]

Chelelgo, G.C., Siriba, D.N. and Biamah, E.K. (2016). Micro Hydro Potential Modelling: Integrating GIS into Energy Alternatives for Climate Change Mitigation. *Journal of Geoscience and Environment Protection*, **4**(08), p.47. doi: 10.4236/gep.2016.48005

Cheng, C., Liu, B., Chau, K.W., Li, G. and Liao, S. (2015). China′ s small hydropower and its dispatching management. *Renewable and Sustainable Energy Reviews*, **42**:43-55. doi: 10.1016/j.rser.2014.09.044

Cole, S.J. and Moore, R.J. (2008). Hydrological modelling using rain gauge-and radar-based estimators of areal rainfall. *Journal of Hydrology*, **358**(3):159-181.doi: 10.1016/j.jhydrol.2008.05.025

Conner, A.M., Francfort, J.E. and Rinehart, B.N. (1998). US hydropower resource assessment final report. US Department of Energy, Idaho Operations Office

Costa-Cabral, M.C. and Burges, S.J. (1994). Digital elevation model networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas. *Water resources research*, **30**(6): 1681-1692. doi: 10.1029/93WR03512

Crépon, O (2009). Re-assessing French hydropower potential. Int. J. Hydropower Dams, 5:47–49

Dawson, C.W., Abrahart, R.J., Shamseldin, A.Y. and Wilby, R.L. (2006). Flood estimation at ungauged sites using artificial neural networks. *Journal of hydrology*, **319**(1):391-409. doi: 10.1016/j.jhydrol.2005.07.032

De Roo, A.P.J., Wesseling, C.G. and Van Deursen, W.P.A. (2000). Physically based river basin modelling within a GIS: the LISFLOOD model. *Hydrological Processes*, **14**(11-12):1981-1992

DIVA-GIS (2016). DIVA-GIS Free spatial data. DIVA-GIS [Available online at http://www.diva-gis.org/Data]

Dowle, M., Srinivasan, A., Gorecki, J., Short, T., Lianoglou, S., Antonyan, E. (2017). data.table: Extension of 'data.frame'

Duchon, J. (1976). Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. Revue française d'automatique, informatique, recherche opérationnelle. *Analyse numérique*, **10**(R3):.5-12

Dudhani, S., Sinha, A.K. and Inamdar, S.S. (2006). Assessment of small hydropower potential using remote sensing data for sustainable development in India. *Energy policy*, **34**(17):3195-3205. doi: 10.1016/j.enpol.2005.06.011

Duncan, N.J. (2014). Mapping Scotland's hydropower resource. PhD Thesis, Edinburgh University

Eddelbuettel, D., Francois, R., Bates, D., Binxiang, Ni. (2017). Rcpp Integration for the 'Armadillo' Templated Linear Algebra Library v 0.7.960.1.2

Eddelbuettel, D. and Francois, R. (2017). Rcpp: Seamless R and C++ Integration 0.12.12

EBS (2016). EBS Pipe Systems: Glass fiber reinforced plastic (GRP) pipes. EBS ECE Pipe Systems, Bakanlıklar, Ankara [Available online at http://www.ebsboru.com/kataloglar/eng/ebs-grp-katalog.pdf]

EHSA (2010). Hydropower and environment. European Small Hydropower Association, Brussels, Belgium [Available online at

https://ec.europa.eu/energy/intelligent/projects/sites/iee-projects/files/projects/documents/sherpa_report_on_environmental_integration.pdf]

EIADCAS (Environment Impact Assessmenet Department, Chinese Academy of Science), RIPYWR (Research Institute for Protection of Yangtze Water Resources) (1995). Environmental Impact Statement for the Yangtze Three Gorges Project (A Brief Edition). Science Press, Beijing (72 :) Environmental impact assessments of the Three Gorges Project in China: Issues and interventions (PDF Download Available). [Available online at

https://www.researchgate.net/publication/260725538_Environmental_impact_assessments_of_the_Three_Gorges_Project_in_China_Issues_and_interventions [accessed Nov 09 2017]]

El Kenawy, A.M., Lopez-Moreno, J.I., McCabe, M.F. and Vicente-Serrano, S.M. (2015). Evaluation of the TMPA-3B42 precipitation product using a high-density rain gauge network over complex terrain in northeastern Iberia. *Global and Planetary Change*, **133**:188-200. doi: 10.1016/j.gloplacha.2015.08.013

Erskine, R.H., Green, T.R., Ramirez, J.A. and MacDonald, L.H. (2006). Comparison of grid-based algorithms for computing upslope contributing area. *Water Resources Research*, **42**(9). doi: 10.1029/2005WR004648

Erskine, R.H., Green, T.R., Ramirez, J.A. and MacDonald, L.H. (2007). Digital elevation accuracy and grid cell size: effects on estimated terrain attributes. *Soil Science Society of America Journal*, **71**(4):.1371-1380. doi:10.2136/sssaj2005.0142

ESRI (1993). Digital Chart of the World. Environmental Systems Research Institute, Redlands, California, USA

ESRI (2015). ArcWorld. Environmental Systems Research Institute, Redlands, California, USA

ESRI (2017). Esri home website. Environmental Systems Research Institute, Redlands, California [Available online at https://www.esri.com/en-us/home]

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L. and Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D. and Alsdorf, D. (2007). The shuttle radar topography mission. *Reviews of geophysics*, **45**(2):1-33. doi: 10.1029/2005RG000183

Fairfield, J. and Leymarie, P. (1991). Drainage networks from grid digital elevation models. *Water resources research*, **27**(5):709-717. doi: 10.1029/90WR02658

Fan, F.Y. (2016). FinCal: Time Value of Money, Time Series Analysis and Computational Finance v0.6-3

Feizizadeh, B. and Haslauer, E.M. (2012). GIS-based procedures of hydropower potential for Tabriz basin, Iran. International Journal, :495-502

Fekete, B.M., Vörösmarty, C.J. and Grabs, W. (2002). High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochemical Cycles*, **16**(3):15-1-15-10. doi: 0.1029/1999GB001254

Fibrolux (2017). Reinforced plastic: Advantages and cost of GRP. Fibrolux, Hartlepool, UK [Available online at http://fibrolux.com/main/knowledge/advantages-cost]

Forrest, N. (2006). Potential for Micro-Hydro Generation in a Scottish Catchment. Msc, The University of Edinburgh.

Forrest, N. (2010). Hydrobot: Remote Surveys of National Hydro Resources. In Proceedings of the International Conference HIDROENERGIA 2010

Francfort, J.E., Rinehart, B.N. and Moore, K.M. (1993). Uniform criteria for US Hydropower Resource Assessment. Hydropower evaluation software status report, USDOE, Washington, DC (United States). doi: 10.2172/10171785

Franke, T. (2012). ppso: Particle Swarm Optimization and Dynamically Dimensioned Search, optionally using parallel computing based on Rmpi v0.9-9994

Freeman, T. G (1991). Calculating catchment area with divergent flow based on a regular grid. Comput. Geosci., 17(3):413 – 422

Freeman, T. (1996). What is imaging RADAR? Jet Propulsion Laboratory. Document available at http://southport. jpl. nasa. gov/desc/imagingradarv3. html.

Geocommunity (2014). GIS Data Depot. Geocommunity [Available online at http://data.geocomm.com/]

Geofabrik (2015). GIS downloads. Geofabrik, GmbH Karlsruhe [Available online at https://www.geofabrik.de/data/download.html]

Gesch, D.B., Verdin, K.L., Greenlee, S.K. (1999). United States Geological Survey GTOPO30 (Global 30 ArcSecond Elevation). EOS, 80(6), 67-70 [Available online at https://lta.cr.usgs.gov/GTOPO30]

Ghajarnia, N., Liaghat, A. and Arasteh, P.D. (2015). Comparison and evaluation of high-resolution precipitation estimation products in Urmia Basin-Iran. *Atmospheric Research*, **158**:50-65. doi: 10.1016/j.atmosres.2015.02.010.

Gjermundsen, T. and Jenssen, L. (2001). Economic risk-and sensitivity analyzes for hydropower projects. Pages 23-28 in Hydropower in the New Millenium: Proceedings of the 4th International Conference on Hydropower Development. Bergen, Norway, June 20-122, 2001

Google (2017). Google Earth for download. Google Inc., California, USA [Available online at https://www.google.com/earth/download/gep/agree.html]

Gordon, J.L. (1981). Estimating hydro station costs. *Water Power and Dam Construction*, **33**:31–33

Gordon, J.L. (1983). Hydropower cost estimates. International water power & dam construction, 35(11):30-37. ISSN 0306-400X.

Gordon, J.L. (2003). Determining ‘‘Ballpark’’costs for a proposed project. Hydro Review, 11(1):37-41

Gordon, J.L. (2008). HYDROHELP. Technical report, OEL HYDROSYS Inc. [Available online at http://www.hydrohelp.ca/eng/home.htm]

Gordon, J. and Noel, C. (1986). Economic Limits of Small and Low-Head Hydro. *Water Power and Dam Construction*, **38**(2):23–26

Gordon, J.L. and Penman, A.C. (1979). Quick estimating techniques for small hydro potential. *J. of Power Dam Constr*, **31**:46-55

GRDC (2017). Global Runoff Data Centre. Federal Institute of Hydrology (Bundesanstalt für Gewässerkunde, Germany)
[Available offline at http://www.bafg.de/GRDC/EN/Home/homepage_node.html]

Gupta, S. and Sharma, K. (2016). Optimization of revenue generated by hydro power plant by Bat Algorithm (BA). In Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on (: 3576-3580). IEEE

Haberlandt, U. (2007). Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology*, **332**(1):144-157. doi: 10.1016/j.jhydrol.2006.06.028

Hall, D.G., Hunt, R.T., Reeves, K.S. and Carroll, G.R. (2003). Estimation of economic parameters of US hydropower resources. EERE Publication and Product Library, Washington, DC (United States)

Hall, D.G. (2007). Virtual Hydropower Prospecting–Searching for Hydropower Gold. *International Journal of Hydropower and Dams*, **14**(INL/JOU-07-12860)

Hamududu, B. and Killingtveit, A. (2012). Assessing climate change impacts on global hydropower. *Energies*, **5**(2):305-322. doi:10.3390/en5020305

Harris, I.P.D.J., Jones, P.D., Osborn, T.J. and Lister, D.H. (2014). Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. *International Journal of Climatology*, **34**(3):623-642. doi: 10.1002/joc.3711

Harrison, G.P. and H.W. Whitington (2001). Impact of climate change on hydropower investment. Pages 257-261 in Hydropower in the New Millenium: Proceedings of the 4th International Conference on Hydropower Development. Bergen, Norway, June 20-122, 2001

Harrison, G.P., Whitington, H.W., and A. R.Wallace (2003). Climate change impacts on financial risk in hydropower projects. *IEEE Transactions on Power Systems* **18**(4):1324-1330.

Harvard University (2016). Maps/GIS. [Available online at https://atg.fas.harvard.edu/mapsgis]

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**(1):97-109

Hijmans, R.J., van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., Racine, E.B., Shortridge, A. (2016a). raster: Geographic Data Analysis and Modeling v2.5-8 [Available online at https://cran.r-project.org/web/packages/raster/index.html]

Hijmans, R.J., Williams, E., Vennes, C. (2016b). geosphere: Spherical Trigonometry v1.5-5.

Hock, R. and Holmgren, B. (2005). A distributed surface energy-balance model for complex topography and its application to Storglaciären, Sweden. *Journal of Glaciology*, **51**(172):25-36.

Hofstra, N., Haylock, M., New, M., Jones, P. and Frei, C. (2008). Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research: Atmospheres*, **113**(D21)

Horritt, M.S. and Bates, P.D. (2001). Effects of spatial resolution on a raster based model of flood flow. *Journal of Hydrology*, **253**(1):239-249. doi: 10.1016/S0022-1694(01)00490-5

Hutchinson, M. (1988). Calculation of hydrologically sound digital elevation models. Third International Symposium on Spatial Data Handling, Columbus, Ohio, International Geographical Union

Hutchinson, M. (1989). A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology*, **106**: 211-232

HydroHELP (2008). HydroHELP website. HYDROSYS Inc. [Available online at http://hydrohelp.ca/eng/home.htm]

Hydrosys Consultants (2008). HydroHELP software. HydroSys consultants [Available online at http://hydrohelp.ca/eng/home.htm]

IEA (2010). Renewable Energy Essentials: Hydropower. International Energy Agency, Paris, France [Available online at

https://www.iea.org/publications/freepublications/publication/hydropower_essentials.pdf]

IEA and NEA (2015), Projected costs of generating electricity 2015 edition. International Energy Agency and Nuclear Energy Agency, Paris and Boulogne-Billancourt, France [Available online at https://www.oecd-nea.org/ndd/pubs/2015/7057-proj-costs-electricity-2015.pdf]

IHA (2017), 2017 Hydropower Status Report. International Hydropower Association, London, United Kingdom [Available online at https://www.hydropower.org/2017-hydropower-status-report]

International Rivers (2012). A Risky Climate for Southern African Hydro. Berkeley, California, USA [Available online at https://www.internationalrivers.org/resources/about-international-rivers-3679]

IPCC (2014). Climate Change (2014): Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

IRENA (2012). Renewable Energy Technologies: Cost Analysis Series – Hydropower. International Renewable Energy Agency, Masdar City, Abu Dhabi, UAE [Available online at http://www.irena.org/Contact/index.aspx?mnu=Pri&PriMenuID=22]

Jain, A., Sudheer, K.P. and Srinivasulu, S. (2004). Identification of physical processes inherent in artificial neural network rainfall runoff models. *Hydrological Processes*, **18**(3):571-581. doi: 10.1002/hyp.5502

Japan International Cooperation Agency (2011). Guideline and Manual for Hydropower Development Vol.1, Conventional Hydropower and Pumped Storage Hydropower. Electric Power Development Co., Ltd. JP Design Co., Ltd, Tokyo, Japan.

Jarvis, A., Rubiano, J., Nelson, A., Farrow, A. and Mulligan M. (2004). Practical use of SRTM data in the tropics: comparisons with digital elevation models generated from cartographic data. Working Document no. 198. Cali, International Centre for Tropical Agriculture (CIAT): 32 pp.

Jarvis, A., Reuter, H.I., Nelson, A. and Guevara, E. (2008). Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org)

Jenson, S.K. and Domingue, J.O. (1988). Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric engineering and remote sensing*, **54**(11): 1593-1600

Jing, C., Shortridge, A., Lin, S. and Wu, J. (2014). Comparison and validation of SRTM and ASTER GDEM for a subtropical landscape in Southeastern China. *International Journal of Digital Earth*, **7**(12): 969-992. doi: 10.1080/17538947.2013.807307

Katiraie-Boroujerdy, P.S., Asanjan, A.A., Hsu, K.L. and Sorooshian, S. (2017). Intercomparison of PERSIANN-CDR and TRMM-3B42V7 precipitation estimates at monthly and daily time scales. *Atmospheric Research*, **193**:36-49

Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P. and Thielen, J. (2016). Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level. *Environmental Modelling & Software*, **75**:68-76. doi: 10.1016/j.envsoft.2015.09.009

Kerr Wood Leidal Associates Ltd. (KWL) (2007). Run-of-River Hydroelectric Resource Assessment for British Columbia; Final Report; BC Hydro, BC Transmission Corporation: Vancouver, BC, Canada

Kervyn, M., Ernst, G.G.J., Goossens, R. and Jacobs, P. (2008). Mapping volcano topography with remote sensing: ASTER vs. SRTM. *International Journal of Remote Sensing*, **29**(22):6515-6538

Kong, Y., Wang, J., Kong, Z., Song, F., Liu, Z. and Wei, C. (2015). Small hydropower in China: The survey and sustainable future. *Renewable and Sustainable Energy Reviews*, **48**:425-433. doi: 10.1016/j.rser.2015.04.036

Krause, P., Boyle, D.P. and Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, **5**:.89-97

Krakauer, N.Y., Pradhanang, S.M., Lakhankar, T. and Jha, A.K. (2013). Evaluating satellite products for precipitation estimation in mountain regions: A case study for Nepal. *Remote Sensing*, **5**(8):4107-4123. doi:10.3390/rs5084107

Kucukali, S. (2011). Risk assessment of river-type hydropower plants using fuzzy logic approach. *Energy Policy*, **39**(10):6683-6688

Kuczera, G. and Parent, E. (1998). Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology*, **211**:69–85

Kuenzer, C., Campbell, I., Roch, M., Leinenkugel, P., Tuan, V.Q. and Dech, S., 2013. Understanding the impact of hydropower developments in the context of upstream–downstream relations in the Mekong river basin. *Sustainability science*, **8**(4):565-584. doi: 10.1007/s11625-012-0195-z

Kundzewicz, Z.W. (2008). Climate change impacts on the hydrological cycle. *Ecohydrology & Hydrobiology*, **8**(2):195-203. doi:10.2478/v10104-009-0015-y

Kusre, B.C., Baruah, D.C., Bordoloi, P.K. and Patra, S.C. (2010). Assessment of hydropower potential using GIS and hydrological modeling technique in Kopili River basin in Assam (India). *Applied Energy*, **87**(1) :298-309. doi: 10.1016/j.apenergy.2009.07.019

Lange, N.T. (1999). New mathematical approaches in hydrological modelling—an application of artificial neural networks. Physics and Chemistry of the Earth, Part B: *Hydrology, Oceans and Atmosphere*, **24**(1-2):31-35. doi:10.1016/S1464-1909(98)00007-0

Larentis, D.G., Collischonn, W., Olivera, F. and Tucci, C.E. (2010). Gis-based procedures for hydropower potential spotting. *Energy*, **35**(10):4237-4243. doi: 10.1016/j.energy.2010.07.014

Lea, N.L. (1992). An aspect driven kinematic routing algorithm, in Overland Flow: Hydraulics and Erosion Mechanics, edited by A. J. Parsons and A.D. Abrahams, Chapman and Hall, New York, 1992

Lehner, B., Verdin, K. and Jarvis, A. (2008). HydroSHEDS: New global hydrography derived from spaceborne elevation data. Eos, Transactions American Geophysical Union, 89(10):93-94 [Available online at http://www.hydrosheds.org/]

Lehner, B., Doll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands (GLWD). Journal of Hydrology 296(1-4): 1-22. Data available at http://www.wwfus.org/science/data/globallakes.cfm

Lima, I.B., Ramos, F.M., Bambace, L.A. and Rosa, R.R. (2008). Methane emissions from large dams as renewable energy resources: a developing nation perspective. Mitigation and Adaptation Strategies for Global Change, 13(2):193-206. doi: 10.1029/2005GB002457

Lin (2011). Influence of Dams on River Ecosystem and Its Countermeasures. Journal of Water Resource and Protection, 3:60-66. doi: 10.4236/jwarp.2011.31007

Lowe, P.R. and Ficke, J.M. (1974). The computation of saturation vapor pressure. Defense Technical Information Center

Luo, Z.Q., Xiao-ming, L., Jia-hong, S., Ya-bin, W. and Wang-ping, L. (2007). Deposit 3D modeling and application. Journal of Central South University of Technology, 14(2):225-229. doi

Ma, Z., Hu, X., Sayer, A.M., Levy, R., Zhang, Q., Xue, Y., Tong, S., Bi, J., Huang, L. and Liu, Y. (2016). Satellite-based spatiotemporal trends in PM2. 5 concentrations: China, 2004–2013. Environmental health perspectives, 124(2):184. doi: 0.1289/ehp.1409481

Majozi, N.P., Mannaerts, C.M., Ramoelo, A., Mathieu, R., Nickless, A. and Verhoef, W. (2017). Analysing surface energy balance closure and partitioning over a semi-arid savanna FLUXNET site in Skukuza, Kruger National Park, South Africa. Hydrology and Earth System Sciences, 21(7):3401

Manning, R. (1891). On the flow of water in open channels and pipes. Transactions of the Institution of Civil Engineers of Ireland. 20:161–207

Marks, D., Domingo, J., Susong, D., Link, T. and Garen, D. (1999). A spatially distributed energy balance snowmelt model for application in mountain basins. *Hydrological Processes*, **13**(12-13):1935-1959. doi: 10.1002/(SICI)1099-1085(199909)13:12/13<1935::AID-HYP868>3.0.CO;2-C

Matthias, H.B, Doujak, E and Angerer, P. (2001). A contribution to ecological-economical aspects of hydro power plants. In: Honningswåg, et al., editors. Hydropower in new millennium. Lisse: Swets and Zeitlinger

Matheron, G. (1960). Krigeage d'un panneau rectangulaire par sa périphérie. Note *géostatistique*, **28**

McMaster, K.J. (2002). Effects of digital elevation model resolution on derived stream network positions. *Water Resources Research*, **38**(4): 131-138. doi: 10.1029/2000WR000150

Mebane, W.R. and Sekhon, J.S. (2015). rgenoud: R Version of GENetic Optimization Using Derivatives v5.7-12.4

Meijer, L (2012). World Hydropower Capacity Evaluation, MSc thesis. Delft University of Technology, Delft, Netherlands

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, **21**(6):1087-1092. doi: 10.1063/1.1699114

METI of Japan and NASA (2011). Advanced Spaceborne Thermal Emission and Relection Radiometer: ASTER. Ministry of Economy, Trade and Industry of Japan (METI) and the United States National Aeronautics and Space Administration (NASA) [Available online at https://asterweb.jpl.nasa.gov/gdem.asp]

Mimikou, M. and E.A. Baltas (1997). Climate change impacts on the reliability of hydroelectric energy production. *Hydrological Sciences Journal*, **42**(5):661-678

Minville, M., Brissette, F., Krau, S. and Leconte, R. (2009). Adaptation to climate change in the management of a Canadian water-resources system exploited for hydropower. *Water Resources Management*, **23**(14):2965-2986

Mispan, M. R., Rasid, M. Z. A., Rahman, N. F. A., Khalid, K., Haron, S. H., & Ahmad, N. (2015). Assessment of ASTER and SRTM Derived Digital Elevation Model for Highland Areas of Peninsular Malaysia Region. *International Research Journal of Engineering and Technology* (IRJET), **2**(6):316-320. ISSN: 2395-0056

Moghadas, S. (2016). Urban Runoff and Snowmelt: Quantity and Quality Processes in Snow Deposits and Hydrologic Abstractions. Doctoral PhD Thesis. Lulea University of Technology.

Moody, J.A. and Troutman, B.M. (2002). Characterization of the spatial variability of channel morphology. *Earth Surface Processes and Landforms*, **27**(12): 1251-1266. doi: 10.1002/esp.403

Monteith, J.L. (1965). Evaporation and environment, In: Proceedings of the 19th Symposium of the Society for Experimental Biology. Cambridge University Press, New York, NY pp. 205–233

Moore, R.J. (1985). The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal*, **30**(2):273-297. doi: 10.1080/02626668509490989

Movable Type Scripts (2015). Calculate distance, bearing and more between Latitude/Longitude points [Available online at http://www.movable-type.co.uk/scripts/latlong.html]

Mu, Q., Zhao, M. and Running, S.W. (2013). Modis Global Terrestrial Evapotranspiration (ET) Product—Algorithm Theoretical Basis Document (Collection 5). [Available online: http://www. ntsg. umt. edu/sites/ntsg. umt. edu/files/MOD16_ATBD. Pdf].

Mueller, B., Seneviratne, S.I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J.B., Guo, Z. and Jung, M. (2011). Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophysical Research Letters*, **38**(6). doi: 10.1029/2010GL046230

Mullen, K.M (2014). Continuous Global Optimization in R. *Journal of Statistical Software*, **60**(6). doi:10.18637/jss.v060.i06

NASA (2017). Global Climate Change, Vital Signs of the Planet. National Aeronautics and Space Administration, Washington DC, USA [Available online at https://climate.nasa.gov/]

NASA (2015). NASA Shuttle Radar Topography Mission (SRTM) Version 3.0 Global 1 arc second Data Released over Asia and Australia. National Aeronautics and Space Administration, Washington DC, USA [Available online at https://earthdata.nasa.gov/nasa-shuttle-radar-topography-mission-srtm-version-3-0-global-1-arc-second-data-released-over-asia-and-australia]

NASA (2017). Goddard Earth sciences data and information services center (GES DISC). National Aeronautics and Space Administration, Washington DC, USA [Available online at https://earthdata.nasa.gov/about/daacs/daac-ges-disc]

NASA/NGA (2003). SRTM Water Body Data Product Specific Guidance, Version 2.0. 4pp. National Aeronautics and Space Administration and National Geospatial-Intelligence Agency. [Available online at https://lta.cr.usgs.gov/srtm_water_body_dataset]

NASA Jet Propulsion Laboratory (2017). ASTER Global Digital Elevation Map Announcement. National Aeronautics and Space Administration, Washington DC, USA [Available online at https://asterweb.jpl.nasa.gov/gdem.asp]

Nash, J.E. and Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, **10**(3):282-290. doi: 10.1016/0022-1694(70)90255-6

Natural Earth (2016). Lakes and reservoirs physical vectors. Natural Earth [Available online at http://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-lakes/]

Natural Resources Canada (2017), RETScreen. [Available online at http://www.nrcan.gc.ca/energy/software-tools/7465]

NCEP CSFR (2016). Climate Forecast System. National Oceanic and Atmospheric Administration National Centers for Environmental Information, Asheville, North Carolina, USA [Available online at https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/climate-forecast-system-version2-cfsv2]

NEA (2016). China's 13th Renewable Energy Development Five Year Plan, National Energy Administration. National Energy Administration, Beijing, China [Available online at http://www.nea.gov.cn/2016-12/19/c_135916140.htm]

Neitsch, S.L., Arnold, J.G., Kiniry, J.R. and Williams, J.R. (2011). Soil and water assessment tool theoretical documentation version 2009 [Available online at http://swat.tamu.edu/media/99192/swat2009-theory.pdf]

NGA (2017). Digital Terrain and Elevation Data. National Geospatial-Intelligence Agency, Springfield, Virginia

NCAR (2016). Climate Data: Precipitation datasets: Overview and comparison table. National Centre for Atmospheric Research, Boulder, Colorado, USA [Available online at https://climatedataguide.ucar.edu/climate-data/precipitation-data-sets-overview-comparison-table]

NTT Data Corporation (2016). AW3D World 3D Topographic Data. Nippon Telegraph and Telephone Corporation. Tokyo, Japan. [Available online at http://www.aw3d.jp/en/]

O'Callaghan, J.F. and Mark, D.M. (1984). The extraction of drainage networks from digital elevation data. *Computer vision, graphics, and image processing*, **28**(3): 323-344. doi: 10.1016/S0734-189X(84)80011-0

Ouerghi, S., ELsheikh, R.F.A., Achour, H. and Bouazi, S. (2015). Evaluation and Validation of Recent Freely-Available ASTER-GDEM V. 2, SRTM V. 4.1 and the DEM Derived from

Topographical Map over SW Grombalia (Test Area) in North East of Tunisia. *Journal of Geographic Information System*, **7**(3):266

Paish, O. (2002). Small hydro power: technology and current status. *Renewable and sustainable energy reviews*, **6**(6):537-556. doi:10.1016/S1364-0321(02)00006-0

Pandey, A., Lalrempuia, D. and Jain, S.K. (2015). Assessment of hydropower potential using spatial technology and SWAT modelling in the Mat River, southern Mizoram, India. *Hydrological Sciences Journal*, **60**(10):1651-1665. doi: 10.1080/02626667.2014.943669

Papantonis, D. (2001). Small hydro power stations. Simeon, Athens, Greece.

Pechlivanidis I.G., McIntyre N., Wheater H.S. (2010). Calibration of the semi-distributed PDM rainfall-runoff model in the Upper Lee catchment, UK, *Journal of Hydrology*, **386** (1-4): 198-209

Pechlivanidis I.G., Jackson, B., McIntyre, N., Wheater, H.S. (2011). Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications, *Global NEST Journal*, **13** (3): 193-214

Peña-Arancibia, J.L., van Dijk, A.I., Renzullo, L.J. and Mulligan, M. (2013). Evaluation of precipitation estimation accuracy in reanalyses, satellite products, and an ensemble method for regions in Australia and South and East Asia. *Journal of Hydrometeorology*, **14**(4):1323-1333. doi: 10.1175/JHM-D-12-0132.1

Penche, C. (2004). Layman's Handbook on How to Develop a Small Hydro Site, CEC DG TREN and ESHA, June 1998

Peng, J., Loew, A., Chen, X., Ma, Y. and Su, Z. (2016). Comparison of satellite-based evapotranspiration estimates over the Tibetan Plateau. *Hydrology and Earth System Sciences*, **20**(8):3167-3182. doi: 10.5194/hess-20-3167-2016

Penman, H.L. (1948). Natural evaporation from open water, bare soil, and grass. *Proceedings of the Royal Society*, London A **193**:120–146

Pierce (2017). ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files v1.16

Þórarinsdóttir, T. (2012). Development of a methodology for estimation of technical hydropower potential in Iceland using high resolution hydrological modelling

Pryde, J.K., Osorio, J., Wolfe, M.L., Heatwole, C., Benham, B. and Cardenas, A. (2007). Comparison of watershed boundaries derived from SRTM and ASTER digital elevation datasets and from a digitized topographic map. American Society of Agricultural and Biological Engineers, Paper Number: 072093

Punys, P., Dumbrauskas, A., Kvaraciejus, A. and Vyciene, G. (2011). Tools for small hydropower plant resource planning and development: A review of technology and applications. **Energies**, *4*(9):1258-1277. doi:10.3390/en4091258

Qi, W., Zhang, C., Fu, G., Sweetapple, C. and Zhou, H. (2016). Evaluation of global fine-resolution precipitation products and their uncertainty quantification in ensemble discharge simulations. *Hydrology and Earth System Sciences*, **20**(2):903-920. doi: 0.5194/hess-20-903-2016

Quan, J., Tie, X., Zhang, Q., Liu, Q., Li, X., Gao, Y. and Zhao, D. (2014). Characteristics of heavy aerosol pollution during the 2012–2013 winter in Beijing, China. *Atmospheric Environment*, **88**:83-89. doi: 10.1016/j.atmosenv.2014.01.058

Rahi, O.P., Chandel, A.K. and Sharma, M.G. (2012). Optimization of hydro power plant design by particle swarm optimization (PSO). *Procedia Engineering*, **30**:418-425

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Ramanathan, V. and Carmichael, G. (2008). Global and regional climate changes due to black carbon. *Nature geoscience*, **1**(4):221-227. doi:10.1038/ngeo156

Rexer, M. and Hirt, C. (2014). Comparison of free high-resolution digital elevation data sets (ASTER GDEM2, SRTM v2. 1/v4. 1) and validation against accurate heights from the Australian National Gravity Database. *Australian Journal of Earth Sciences*, **61**(2):213-226.

Reanalyses.org (2016). Reanalyses.org Home Page Advancing Reanalysis [Available online at https://reanalyses.org/]

RETScreen (2004). RETScreen Clean Energy Management Software system. RETScreen International, Natural Resources Canada, Government of Canada [Available online at http://www.nrcan.gc.ca/energy/software-tools/7465]

RETScreen (2005). Small Hydro Project Analysis within Clean Energy Project Analysis RETScreen Engineering & Cases Textbook Third Edition. RETScreen International, Natural

Resources Canada, Government of Canada [Available online at http://publications.gc.ca/collections/collection_2007/nrcan-rncan/M154-13-2005E.pdf]

Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.K. and Bloom, S. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of climate*, **24**(14):3624-3648. doi: 10.1175/JCLI-D-11-00015.1

Roe, G.H., Baker, M.B. and Herla, F. (2017). Centennial glacier retreat as categorical evidence of regional climate change. *Nature Geoscience*, **10**(2):95-99. doi:10.1038/ngeo2863

Rusli, N., Majid, M.R. and Din, A.H.M. (2014). Google Earth's derived digital elevation model: A comparative assessment with Aster and SRTM data. In IOP Conference Series: *Earth and Environmental Science* **18**, No. 1, p. 012065). IOP Publishing

Samaniego, L., Kumar, R. and Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, **46**(5). doi: 10.1029/2008WR007327

Sample, J.E., Duncan, N., Ferguson, M. and Cooksley, S. (2015). Scotland′ s hydropower: Current capacity, future potential and the possible impacts of climate change. *Renewable and Sustainable Energy Reviews*, **52**:111-122

Schaefli, B., Hingray, B. and Musy, A. (2007). Climate change and hydropower production in the Swiss Alps: quantification of potential impacts and related modelling uncertainties. *Hydrology and Earth System Sciences Discussions*, **11**(3):1191-1205

Shah, R. and Mishra, V. (2014). Evaluation of the reanalysis products for the monsoon season droughts in India. *Journal of Hydrometeorology*, **15**(4):1575-1591. doi:10.1175/JHM-D-13-0103.1

Sharma, R.N., Chand, N., Sharma, V. and Yadav, D. (2015). Decision support system for operation, scheduling and optimization of hydro power plant in Jammu and Kashmir region. *Renewable and Sustainable Energy Reviews*, **43**:1099-1113

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In Proceedings of the 1968 23rd ACM national conference (pp. 517-524). ACM. doi: 10.1145/800186.810616

Singal, S.K. and Saini, R.P. (2008). Cost analysis of low-head dam-toe small hydropower plants based on number of generating units. *Energy for Sustainable development*, **12**(3):55-60. doi: 10.1016/S0973-0826(08)60439-1

Singal, S.K. and Saini, R.P. (2007). Analytical approach for cost estimation of low head small hydro power schemes. In International Conference on Small Hydropower-Hydro Sri Lanka, 22: 24

Singal, S.K., Saini, R.P. and Raghuvanshi, C.S. (2010). Analysis for cost estimation of low head run-of-river small hydropower schemes. *Energy for sustainable Development*, **14**(2):117-126. doi: 10.1016/j.esd.2010.04.001

Singh, S. and Singh, S. (1995). Quantitative analysis of watershed geomorphology using remote sensing techniques. *Annals of Arid Zone*, **34**:243-252

Snowdonia Hydro (2014). Snowdonia Hydro Home Website. [Available online at http://snowdoniahydro.com/micro-hydro/hydro-faqs/]

Snow Hydrology (1956). Snow hydrology; summary report of the snow investigations. North Pacific Division, Corps of Engineers, U.S. Army (1956)

Soetaert, K. (2016). rootSolve: Nonlinear Root Finding, Equilibrium and Steady-State Analysis of Ordinary Differential Equations v1.7.

Soulis, K.X., Manolakos, D., Anagnostopoulos, J. and Papantonis, D. (2016). Development of a geo-information system embedding a spatially distributed hydrological model for the preliminary assessment of the hydropower potential of historical hydro sites in poorly gauged areas. *Renewable Energy*, **92**:222-232. doi: 10.1016/j.renene.2016.02.013

Sovacool, B.K., Nugent, D. and Gilbert, A. (2014). Construction cost overruns and electricity infrastructure: an unavoidable risk? *The Electricity Journal*, **27**(4):112-120

Sovacool, B.K., Kryman, M. and Laine, E. (2015). Profiling technological failure and disaster in the energy sector: A comparative analysis of historical energy accidents. *Energy*, **90**:2016-2027

Statista (2017). Global electricity prices by select countries in 2015 (in U.S. dollar cents per kilowatt hour). Statista, New York, USA [Available online at https://www.statista.com/statistics/263492/electricity-prices-in-selected-countries/]

Storn, R. and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, **11**(4):341-359

Suarez, M.J., Rienecker, M.M., Todling, R., Bacmeister, J., Takacs, L., Liu, H.C., Gu, W., Sienkiewicz, M., Koster, R.D., Gelaro, R. and Stajner, I. (2008). The GEOS-5 Data Assimilation System-Documentation of Versions 5.0. 1, 5.1. 0, and 5.2. 0. National Aeronautics and Space Administration, Washington DC, USA

Tan, M.L., Ibrahim, A.L., Duan, Z., Cracknell, A.P. and Chaplot, V. (2015). Evaluation of six high-resolution satellite and ground-based precipitation products over Malaysia. *Remote Sensing*, **7**(2):1504-1528. doi:10.3390/rs70201504

Tarboton, D. G. (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models, *Water Resources Research* **33**(2), pp. 309–319. doi: 10.1029/96wr03137

Tarboton, D.G., Chowdhury, T.G. and Jackson, T.H.(1994). A spatially distributed energy balance snowmelt model. Utah Water Research Laboratory.

Thiessen, A.H. (1911). Precipitation averages for large areas. *Monthly weather review*, **39**(7):1082-1089. doi: 10.1175/1520-0493(1911)39<1082b:PAFLA>2.0.CO;2

Trading economics (2016), Trading economics labour costs [Available online at https://tradingeconomics.com/country-list/labour-costs]

UKForex (2016), Historical Exchange Rates. London, United Kingdom [Available online at https://www.ofx.com/en-gb/about-us/]

UNIDO (2016), World Small Hydropower Development Report 2016. United Nations Industrial Development Organization, Vienna and International Center on Small Hydro Power, Hangzhou [Available online at: https://www.unido.org/fileadmin/user_media_upgrade/What_we_do/Topics/Energy_access/WSHPDR_Executive_Summary_2016.pdf]

Urbanek, S. (2012). proj4: A simple interface to the PROJ.4 cartographic projections library v 1.0-8.

USGS (2016). Landsat Mission Data. United States Geological Survey, Sioux Falls, South Dakota [Available online at https://landsat.usgs.gov/]

United Nations (2003). Water for People, Water for Life – the United Nations World Water Development Report. United Nations.

University of East Anglia Climatic Research Unit (2013). Climatic Research Unit Home Page. University of East Anglia, Norwich, UK [Available online at http://www.cru.uea.ac.uk/]

Utter Pradesh Power Transmission Corporation Limited (2016). Current Carrying Capacity of Overhead Transmission Line ACSR Conductor. Uttar Pradesh Power Transmission Corporation Limitied, Uttar Pradesh, India [Available online at http://upptcl.org/tech_info/current_carrying.htm]

Valipour, M., Sefidkouhi, M.A.G. and Raeini, M. (2017). Selecting the best model to estimate potential evapotranspiration with respect to climate change and magnitudes of extreme events. *Agricultural Water Management*, **180**:50-60. doi: 10.1016/j.agwat.2016.08.025

van Etten, J. (2017). gdistance: Distances and Routes on Geographical Grids v 1.2-1.

Vrugt, J.A. (2003). A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, **39**(8). ISSN 0043-1397. doi:10.1029/2002WR001642

Vu, M.T., Raghavan, S.V. and Liong, S.Y. (2012). SWAT use of gridded observations for simulating runoff-a Vietnam river basin study. *Hydrology and Earth System Sciences*, **16**(8), p.2801. doi:10.5194/hess-16-2801-2012

Wallace (2018). Professor Wallace expert opinion. Edinburgh University, Edinburgh.

Wang K. (2014). Sampling Biases in Datasets of Historical Mean Air Temperature over Land. *Scientific Reports*, **4:**4637. doi: 10.1038/srep04637

Wang, B., Liang, X.J., Zhang, H., Wang, L. and Wei, Y.M. (2014). Vulnerability of hydropower generation to climate change in China: Results based on Grey forecasting model. *Energy Policy*, **65**:701-707. doi:10.1016/j.enpol.2013.10.002

Wang, W., Yang, X. and Yao, T. (2012). Evaluation of ASTER GDEM and SRTM and their suitability in hydraulic modelling of a glacial lake outburst flood in southeast Tibet. *Hydrological Processes*, **26**(2):213-225. doi: 10.1002/hyp.8127

Warmerdam (2000). Geospatial Data Abstraction Library. [Available online at http://www.gdal.org/]

Wickham, H. (2016). plyr: Tools for Splitting, Applying and Combining Data v1.8.4

Willmott, C.J. (1984). On the evaluation of model performance in physical geography. *Spatial statistics and models*, 443-460.

Willmott, C. J., Rowe, C.M., and Philpot, W. D. (1985). Small-scale climate maps: a sensitivity analysis of some common assumptions associated with grid point interpolation and contouring, *American Cartographer*, **12**:5-16

Wilson, E.M. (2000). The International Energy Agency – Implementing agreement for hydropower technologies and programmes: Assesment methods for small-hydro projects. International Energy Agency, Paris France [Available online at http://www.ieahydro.org/media/c617f800/Assessment%20Methods%20for%20Small-Hydro%20Projects.pdf]

Wisser, D., Fekete, B.M., Vörösmarty, C.J. and Schumann, A.H. (2010). Reconstructing 20th century global hydrography: a contribution to the Global Terrestrial Network-Hydrology (GTN-H). *Hydrology and Earth System Sciences*, **14**(1):1-24. doi: 10.5194/hess-14-1-2010

World Bank (2017). World Bank Open Data, World Bank. Washington DC, USA [Available online at https://data.worldbank.org/]

World Energy Council (2010). World Energy Resources: 2010 Survey. London, UK [Available online at https://www.worldenergy.org/publications/2010/survey-of-energy-resources-2010/]

World Energy Council (2013). World Energy Perspective: Cost of Energy Technologies. London, UK [Available online at

https://www.worldenergy.org/wp-content/uploads/2013/09/WEC_J1143_CostofTECHNOLOGIES_021013_WEB_Final.pdf]

World Energy Council (2017). Energy Resources – Hydropower. London, UK [Available online at https://www.worldenergy.org/data/resources/resource/hydropower/]

WWF (2004). Hydropower and the environment: Towards better decision making. World-Wide Fund for Nature. Wolrd-Wide Fund for Nature, Gland, Vaud, Switzerland. [Available online at http://www.un.org/esa/sustdev/sdissues/energy/op/hydro_collier.pdf]

WWF Global (2017). Yangtze Basin lakes shrinking as climate change, development takes its toll. World-Wide Fund for Nature, Gland, Vaud, Switzerland. [Available online at http://wwf.panda.org/wwf_news/?203195/Yangtze-Basin-lakes-shrinking-as-climate-change-development-takes-its-toll]

Xu, X., Tan, Y. and Yang, G. (2013). Environmental impact assessments of the Three Gorges Project in China: Issues and interventions. *Earth-Science Reviews,* **124**:115-125. doi: 10.1016/j.earscirev.2013.05.007

Yang, W., Ferraro, R.R., Arkin, P.A. and Wick, G.A. (2015). Comparison of Global Atmospheric Rivers Depicted from Satellite and NWP Reanalysis. In AGU Fall Meeting Abstracts

Yasutomi, N., Hamada, A. and Yatagai, A. (2011). Development of a long-term daily gridded temperature dataset and its application to rain/snow discrimination of daily precipitation. *Global Environ. Res*, **15(**2):165-172

Yatagai, A., Arakawa, O., Kamiguchi, K., Kawamoto, H., Nodzu, M.I. and Hamada, A. (2009). A 44-year daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Sola*, **5**:137-140. doi:10.2151/sola.2009-035

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N. and Kitoh, A. (2012). APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bulletin of the American Meteorological Society*, **93(**9), pp.1401-1415. doi:10.1175/BAMS-D-11-00122.1

Yin, H., Donat, M.G., Alexander, L.V. and Sun, Y. (2015). Multi-dataset comparison of gridded observed temperature and precipitation extremes over China. *International Journal of Climatology*, **35(**10):2809-2827. doi: 10.1002/joc.4174

Yuce, M.I. and Yuce, S. (2016). Pre-feasibility Assessment of Small Hydropower Projects in Turkey by RETScreen. *Journal-American Water Works Association*, **108(**5):E269-E275. doi: 10.5942/jawwa.2016.108.0046

Zhang, K. (2010) Remote Sensing (RS) GIMMS NDVI Based Daily ET and Monthly PET for Continental US (CONUS) from 1983 to 2006. from Numerical Terradynamic Simulation Group. University of Montana ftp://ftp. ntsg. umt. edu/pub/data/CONUS_ET

Zhang, Q.F., Smith, B. and Zhang, W. (2012). Small hydropower cost reference model. Oak Ridge National Laboratory. US Department Of Energy

Zhang, X., Dong, Z., Gupta, H., Wu, G. and Li, D. (2016a). Impact of the Three Gorges Dam on the Hydrology and Ecology of the Yangtze River. *Water,* **8(**12) p.590. doi:10.3390/w8120590

Zhang, Y., Peña-Arancibia, J.L., McVicar, T.R., Chiew, F.H., Vaze, J., Liu, C., Lu, X., Zheng, H., Wang, Y., Liu, Y.Y. and Miralles, D.G. (2016b). Multi-decadal trends in global terrestrial evapotranspiration and its components. *Scientific reports*, **6**, p.19124. doi: 10.1038/srep19124

Zhou, Y., Hejazi, M., Smith, S., Edmonds, J., Li, H., Clarke, L., Calvin, K. and Thomson, A. (2015). A comprehensive view of global potential for hydro-generated electricity. *Energy & Environmental Science*, **8**(9):2622-2633. doi: 10.1039/C5EE00888C